

## 데이터 분석 최종결과보고서

### I. 참가자 정보

제 목	어린이, 노약자를 위한 횡단보도 클러스터링 및 스마트 횡단보도 최적 입지 선정	
팀 명	슈터디	
성 명	손지영	
연락처	휴대폰	010-8514-9384
	E-mail	sgsgk1213@naver.com

# 데이터 분석 최종결과보고서

어린이, 노약자를 위한 횡단보도 클러스터링  
및 스마트 횡단보도 최적 입지 선정

슈터디
김아진
손지영
원지수
최세환

## 목차

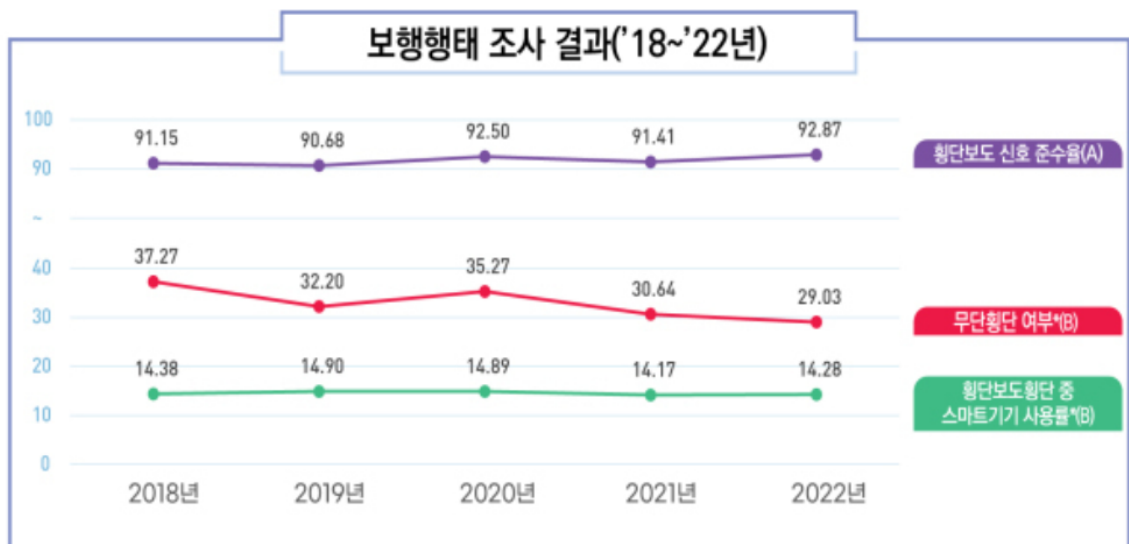
<b>1. 개요</b>	4p
1.a 선정 배경 및 동기	4p
1.b 연구의 필요성	6p
1.c 분석/시각화 결과 내용 요약	6p
<b>2. 연구</b>	8p
2.a 관련연구 및 해결하고자 하는 문제	8p
<b>3. 데이터</b>	8p
3.a 사용한 데이터 소개	8p
<b>4. 전처리</b>	9p
4.a 신고데이터 종합	9p
4.b 횡단보도별 사고데이터 집계	10p
4.c 횡단보도별 인근 학교, 어린이집, 노인시설 개수 데이터 추가	11p
4.d 횡단보도별 인근 버스정류장 개수 데이터 추가	11p
4.e 횡단보도별 인근 아파트 세대수, 단지수 데이터 추가	11p
4.f 횡단보도별 인근 무인단속카메라 개수 데이터 추가	12p
4.g 최종 데이터	13p
<b>5. 분석/시각화 결과 상세내용(EDA)</b>	14p
5.a map	14p
5.b plot	19p
<b>6. 모델링 &amp; 결과 해석</b>	21p
6.a 사고다발지역 예측 모델	21p
6.b 피쳐 중요도 추출	23p
6.c 클러스터링	24p
6.d 피쳐 중요도를 활용한 가중치 계산	30p
<b>7. 결론</b>	31p
7.a 최적입지 선정	31p
7.b 한계점 및 향후 발전방향	40p
7.c 연구를 통한 기대효과	41p
<b>8. 참고 문헌 및 데이터 출처</b>	43p

## 1. 개요

보고서에 앞서, 본 연구에 진행된 모든 전처리, 모델링 등의 자세한 코드 파일은 ([https://colab.research.google.com/drive/1MCNFzRjk7vxOK-zYAezduxif5SJ\\_HW0G?usp=sharing](https://colab.research.google.com/drive/1MCNFzRjk7vxOK-zYAezduxif5SJ_HW0G?usp=sharing))에서 확인할 수 있다.

### a. 선정 배경 및 동기

대한민국의 교통사고 보행 중 사망자수는 OECD 회원국 평균의 3배이다. 교통사고를 줄이기 위해서는 법안 개정 뿐만이 아닌 물리적인 교통시설물의 설치가 절실히 필요하다. 예로, 충남경찰청의 개선 사업 추진(검증된 교통시설물 설치)으로 해당 지점의 교통사고가 15~24% 감소한 것으로 나타났다. 또한 행정안전부가 교통사고 잦은 곳을 대상으로 개선사업을 시행한 결과 교통사고 사망자수는 80%, 교통사고 건수는 30.3% 감소한 결과를 보였다. 신호기 증설, 무단횡단 방지펜스 설치 등 간단한 교통 안전시설 개선만으로 효과가 매우 높은 것으로 나타난다. 허언욱 행정안전부 안전정책실장은 “교통사고가 자주 발생하는 곳에 대해 사고 원인 분석에 따른 맞춤형 개선대책이 무엇보다 중요하다.”고 언급한 바 있다.



[그래프 1] 최근 5년간 국민들의 보행행태의 조사 결과 (국토교통부 제공)

위 [그래프 1]을 보면 횡단보도 신호 준수율은 높은 편이나, 여전히 무단횡단이 30%에 육박한다는 점, 그리고 횡단보도 횡단 중 스마트기기 사용률이 꾸준한 점을 알 수 있다. 이와 더불어 전체 보행자 사망사고의 23.1%가 횡단보도 내에서

발생한다는 조사 결과가 존재한다. 따라서, 횡단보도 근방 사고를 줄이기 위해 운전자와 보행자 모두에게 도움이 되는 ‘스마트 횡단보도’ 도입을 제안한다.

‘스마트 횡단보도’란 보행자 및 운전자의 특성을 고려하여 다양한 시청각 시설물을 통해 보행안전 경각심을 높이고 사고를 방지하는 시설물을 의미한다 (Seoul, 2016). 보행자의 움직임에 따라 조명이 이동함으로써 운전자가 보행자를 더 잘 인지하게 하는 집중조명, 횡단보도에 대한 보행자의 인지를 효과적으로 환기하는 바닥신호등, 로고라이트 및 음성안내기, 횡단보도에 대한 운전자의 인지 정도를 높이는 정지선 라이트가 그 예이다. 아래의 [사진 1]처럼 스마트 횡단보도는 보행자가 더욱 안전한 보행환경을 누릴 수 있도록 돕는 시설이라고 할 수 있다.



[사진 1] 스마트 횡단보도 (출처: 스마트서울 포털)

충남 천안시는 2022년 ‘스마트 교통도시 천안’ 기틀 원년의 해로 삼고 주요 간선도로 교차로 교통흐름 개선 및 보행자 교통사고 예방에 총력을 기울였다. 그에 따라 ‘스마트 횡단보도’를 설치하였으며, 올해도 계속 이어나갈 예정이다. [사진 1]과 같이 인공지능(AI) 기반 지능형 CCTV를 통해 보행자와 차량을 상시 감시하며 상충 위험도를 자동으로 인식하고 이를 전광판에 표출 및 음성안내를 제공한다. 교통 약자 보행 시에는 보행신호를 연장하거나 보행자의 대기 시간도 줄일 수 있어 보행자의 안전 확보는 물론 교통정체를 해소하는 효과도 기대해볼 수 있다.

## b. 연구의 필요성

횡단보도에서 일어나는 교통사고를 방지하고자, 보행자 보호의무 강화 도로교통법이 개정되었다. 하지만 개정 이후에도 도로 위 상황은 크게 바뀌지 않았다. 이는 단순 법 개정뿐만 아니라, 다른 해결 방안이 필요하다는 것을 의미한다.

정부는 이미 2020년 '한국판 뉴딜'에서 '국민안전 SOC 디지털화'를 10대 추진 과제로 선정하였다. 이는 SOC에 해당하는 교통에 인공지능·사물인터넷·5G 등 디지털 기술을 입히는 것이며, 차세대지능형교통시스템(C-ITS)를 구축을 목표로 한다. 이를 토대로 서울특별시 성동구와 양천구를 '스마트시티 특구'로 지정하였는데, 성동구는 보행량이 많아 교통사고 위험이 높은 14개 횡단보도에 '스마트 횡단보도'를 구축해 차량정지선 위반건수를 70% 가까이 줄일 수 있었다. 또한 보행자 교통사고도 21.5% 감소 및 사망자 0명이라는 효과를 보였다.

하지만 '스마트 횡단보도' 도입에 필요한 객관적인 근거와 예산이 부족하다는 한계점이 존재한다. 이를 해결하기 위해서는 '스마트 횡단보도' 도입이 필요한 최적 입지 선정을 위한 객관적인 지표가 필요하다.

## c. 분석/시각화 결과 내용 요약

먼저, 지도 위에 사고, 인구, 연령, 인구 1인당 사고 수, 연령대별 1인당 사고 수, 학교데이터의 시각화를 통하여 횡단보도데이터가 존재하는 지역들에 대한 스마트횡단보도 우선순위 선정이 유효성이 있음을 검증하였다.

둘째로 추가변수들에 대한 시각화를 진행하여 변수의 지역별 편차가 존재함을 박스 플롯과 막대그래프를 통해 확인하였으며, 이에 따라 추가변수들과 기존 변수들의 조합을 통하여 모델링을 진행하였다.

셋째, 수집된 데이터들을 종합하고 이를 분류모델을 거쳐 스마트 횡단보도가 필요한 지역들을 예측하여 도출해내는 모델을 구현하였으나, 해당 모델이 실제 스마트횡단보도를 설치하기위한 지표로서 활용되기에는 부족하다고 판단,

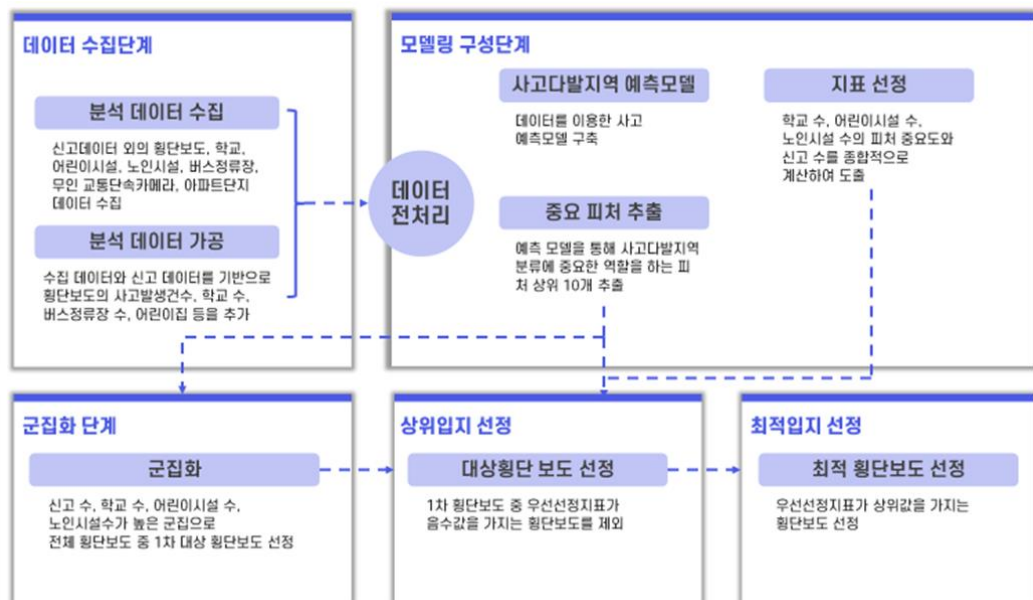
모델에서의 피쳐별 중요도를 도출하고 이를 이용한 군집화를 진행하여 최적입지를 선택하는 것이 바람직하다는 결론을 얻을 수 있었다.

넷째, 앞선 결론에 따라 k-means 알고리즘을 이용하여 횡단보도들에 대한 군집화를 진행하였다. 8개의 군집으로 군집화를 진행하여 분석한 결과 학교, 어린이, 노인시설 등이 많이 인접하여 있으면서, 사고수가 많은 군집으로 횡단보도를 1차 선정하였다.

다섯째, 군집화를 통해 도출된 군집내에서 학교 수와 어린이시설 수, 노인시설 수 피쳐 각각에 피쳐 중요도를 곱하고 합하여 시설 종합 가중치를 도출하였고, 이를 신고 수와 연관하여 최종적인 '우선 선정 평가 지표'를 도출하였다. 도출된 지표에서 음수 값을 가지는 횡단보도들을 제외하고 603개의 스마트 횡단보도 우선 설치 지역을 2차로 도출하였다. 최종적으로 '우선 선정 지표'에서 높은 값을 가지는 횡단보도를 스마트 횡단보도 우선 입지로 판단하였다.

마지막으로 선정 지표를 통해 우선적으로 선택된 횡단보도의 위치데이터를 이용하여 횡단보도들의 일부를 로드뷰를 통해 검토하였으며, 그 결과 선정된 횡단보도들이 통행량과 주변시설들의 분포가 스마트횡단보도의 설치 의의와 합치함을 확인할 수 있었다.

분석 프로세스 개요는 다음과 같다.



## 2. 연구

### a. 관련연구 및 해결하고자 하는 문제

이수현, 서용원, 김세인, 이재경, 윤원주 (2022)는 머신러닝을 활용한 어린이 스마트 횡단보도 최적입지 선정에 연구하였다. 이 연구에서 어린이 이용 시설과의 접근성, 인구수 등의 변수를 포함하여 어린이 교통사고 예측모형을 개발하였고, 해당 모형을 통해 어린이 교통사고에 영향을 미치는 변수들을 선정하고 그에 따른 가중치를 부여하여 최종 입지지수를 산출하였다. K-means clustering을 통해 최적입지를 유형화하였다는 점에서 의의가 있지만, '스마트 횡단보도'를 도입함에 있어 어린이 교통사고에 한정했다는 한계점이 존재한다.

따라서, 본 연구에서는 '스마트 횡단보도'를 도입함에 있어 각 지역별 인구수와 인근 어린이 시설, 노인 및 장애인 시설, 학교, 버스정류장, 횡단보도 근방 사고 횟수 등을 모두 고려하였다. 교통사고 피해자를 어린이로 한정짓지 않고 어린이를 포함하여 노인과 장애인 등 취약계층을 중심으로 보행자 교통사고 피해를 줄이고자 한다. 따라서 사고다발지역 예측 모델을 통해 특정 위치에 횡단보도를 도입할 시 사고 다발 지역 여부를 사전에 고려할 수 있으며, 각 횡단보도에 대한 사고수와 주변 시설과의 가중치 계산을 통해 '스마트 횡단보도' 도입 유무를 결정할 수 있는 '우선 선정 지표'를 제공하고자 한다.

## 3. 데이터

### a. 사용한 데이터 소개

공모전에서 제공한 KP2020, KP2021, NPA2020 데이터와 더불어 본 연구를 위해 필요한 데이터와 더욱 정확한 분석을 위해 수집한 데이터들은 다음과 같다.

더불어 전국횡단보도표준데이터에서 대전광역시, 세종특별자치시, 충청남도(천안시, 서산시, 태안군, 예산군, 서천군 에 한정함)을 구할 수 있었고, 추가로 충청남도 보령시 데이터를 수집하였다.

사용 데이터
전국횡단보도표준데이터
전국노인장애인보호구역표준데이터
전국어린이보호구역표준데이터
전국 버스 정류장 위치 정보
전국무인교통단속카메라표준데이터
k-apt관리비공개의무단지 기본정보
행정구역별 지목별 국토이용현황
연령별 인구현황
보행자 사고다발지역정보 데이터셋

위 데이터들을 사용하여 EDA, 분석, 모델링 등에 사용한 csv 파일들은 다음과 같다. 보고서 외 모든 파일 및 코드들은 추가 제출하였으므로, 자세한 확인이 필요할 시 이를 참고하면 된다.

- i. 횡단보도 외 교통사고 데이터: 횡단보도사건제외신고\_all.csv
- ii. KP2020, KP2021, NPA2020의 신고 데이터에서 횡단보도번호를 제외하고, 중복을 제거한 고유한 횡단보도 사고 데이터: df\_acc\_report\_중복제거\_all.csv
- iii. KP2020, KP2021, NPA2020의 신고 데이터에서 횡단보도번호를 포함하고, 중복을 제거하지 않은 횡단보도 사고 데이터: df\_acc\_report\_all.csv
- iv. 각 횡단보도별 사고 횟수를 포함한 횡단보도사고 데이터: df\_acc

## 4. 전처리

### a. 신고데이터 종합

데이터 분석을 위하여 제공된 npa\_2020, KP2020, KP2021의 통합을 진행하였다.

먼저, 데이터를 통합하기 위하여 위와 같이 주어진 npa\_2020의 데이터 중, kp데이터와 다른형식으로 지정되어 있는 시간 데이터열인 RECV\_CPLT\_DT와 RECV\_CPLT\_TM 열을 이용하여 하나의 datetime object로서 변환하였다.

이후, kp2020, 2021 데이터 또한 시간 데이터인 RECV\_DEPT\_NM열을 datetime object로 변환하여 이를 병합하였다.

최종적으로 병합된 데이터에서 교통사고 데이터를 분석하기 위하여 신고코드를 이용하여 충남, 대전, 세종에서 발생한 신고데이터를 도출하였으며, 사고발생주소를 이용하여 신고의 시도명과 시군구명을 부여하였다.

변환되기 이전의 NPA\_2020데이터와 kp2020 데이터, 최종데이터의 데이터 예시는 다음과 같다.

	RECV_CPLT_DT	RECV_CPLT_TM	NPA_CL	EVT_STAT_CD	EVT_CL_CD	RPTER_SEX		HPPN_OLD_ADDR	HPPN_X	HPPN_Y	SME_EVT_YN
0	20200101	7	13	10	501	2	대전광역시 중구 목동(행정:목동)	360	127.409270	36.333010	Y
1	20200101	132	13	10	501	1	대전광역시 중구 대흥동(대흥동)	499-1	127.421295	36.325575	NaN
2	20200101	39	13	10	501	1		NaN	127.404663	36.341685	NaN

[사진 2]변환되기 이전의 NPA데이터

	RECV_DEPT_NM	RECV_CPLT_DM	NPA_CL	EVT_STAT_CD	EVT_CL_CD	RPTER_SEX		HPPN_PNU_ADDR	HPPN_X	HPPN_Y	SME_EVT_YN
4	충남청	20/12/01 08:17:50.000000000	19	10	401	2.0	충청남도 천안시 서북구 성정동(행정:성정2동)	1259	127.137160	36.826718	NaN
7	대전청	20/12/01 03:24:26.000000000	13	10	406	1.0	대전광역시 유성구 봉명동(행정:온천1동)	575	127.341551	36.353696	NaN
8	충남청	20/12/01 07:37:18.000000000	19	10	402	1.0	충청남도 천안시 서북구 성환읍 성월리(행정:성환읍)	150	127.136142	36.917900	NaN
9	대전청	20/12/01 07:08:44.000000000	13	10	401	1.0	대전광역시 중구 석교동(석교동)	85-14	127.447545	36.306724	NaN
10	대전청	20/12/01 07:09:20.000000000	13	10	401	2.0	대전광역시 중구 석교동(행정:석교동)	88-11	127.447175	36.306694	NaN
...	...	...	...	...	...	...	...	...	...	...	...

[사진 3]변환되기 이전의 KP데이터

	NPA_CL	EVT_STAT_CD	EVT_CL_CD	RPTER_SEX	HPPN_X	HPPN_Y	SME_EVT_YN	date_time	HPPN_ADDR	시도명	시군구명
140996	31	10	406	1	127.298872	36.489101	NaN	2020-12-01 00:20:53	세종특별자치시 세종시 소담동(소담동)	286-3	세종특별자치시 세종시
141007	31	10	403	3	127.250007	36.477555	NaN	2020-12-01 15:02:36	세종특별자치시 세종시 한솔동(한솔동)	1045	세종특별자치시 세종시
141025	31	10	401	3	127.303180	36.601510	N	2020-12-01 13:22:45	세종특별자치시 조치원읍 원리 13		세종특별자치시 조치원읍
141045	31	10	401	1	127.304611	36.603134	NaN	2020-12-01 10:45:45	세종특별자치시 세종시 조치원읍 상리(조치원읍)	39-6	세종특별자치시 세종시

[사진 4]병합되고 필터링 된 최종데이터

## b. 횡단보도별 사고데이터 집계

스마트 횡단보도 최적 입지 분석을 위하여 앞서 병합된 파일에 공공데이터포털과 충청남도 데이터 포털을 통하여 아래와 같은 열을 가지는 대전광역시, 세종특별자치시, 충청남도 일부지역(천안, 선산, 보령, 서천, 태안, 예산)의 횡단보도 위치정보데이터를 수집하였으며, 수집된 데이터의 열은 아래 표와 같다.

이에 대한 위치정보와 신고파일의 위치정보를 통해 haversine 공식을 이용하여 횡단보도의 반경 50m내에서 발생한 교통사고의 개수를 변수로서 파일의 acc\_cnt열에 추가하였다.

	시도 명	시군 구명	도로명	소재지도로 명주소	소재지지번주소	횡단보도 관리번호	횡단보 도종류	자전거횡단 도겸용여부	고원식적 용여부	위도	...	보도턱낮 출여부	점자블 록유무	집중조명 시설유무	관리기관 명	관리기관전 화번호	데이터기 준일자	acc_cnt
0	충청 남도	서산 시	안건로	NaN	충청남도 서산시 갈산 동 163-3	s1	1	N	N	36.795415	...	N	N	N	충청남도 서산시청	041-660-3173	2022-11-16	1
1	충청 남도	서산 시	강경별 로	NaN	충청남도 서산시 부석 면 위평리 665-3	s2	1	N	N	36.720521	...	N	N	N	충청남도 서산시청	041-660-3173	2022-11-16	1
2	충청 남도	서산 시	고운로	NaN	충청남도 서산시 동문 동 53-3	s3	1	N	N	36.784331	...	Y	Y	N	충청남도 서산시청	041-660-3173	2022-11-16	28
3	충청 남도	서산 시	대지제 2로	NaN	충청남도 서산시 읍내 동 17-25	s4	1	N	N	36.776123	...	Y	Y	N	충청남도 서산시청	041-660-3173	2022-11-16	14
4	충청 남도	서산 시	호수공 원1로	NaN	충청남도 서산시 예천 동 38-3	s5	1	N	N	36.774998	...	Y	Y	N	충청남도 서산시청	041-660-3173	2022-11-16	11

[사진 5] acc\_cnt 추가 데이터 예시

### c. 횡단보도별 인근 학교, 어린이집, 노인시설 개수 데이터 추가

다음으로 주제의 목적인 어린이, 노인 보호를 위한 스마트 신호등 설치를 데이터를 이용한 예측, 군집 모델에 반영시키기 위하여 공공데이터포털에서 전구초중고위치정보데이터, 전국어린이집표준데이터, 노인장애인보호구역데이터를 활용하여 1km이내에 있는 학교수 데이터를 sch\_cnt열에 어린이집수 데이터를 ch\_cnt열에 노인정수 데이터를 old\_cnt열에 추가하였다.

시도 명	시군 구명	도로명	소재지도로 명주소	소재지지번주소	횡단보도 관리번호	횡단보 도종류	자전거횡단 도겸용여부	고원식적 용여부	위도	...	보도턱낮 출여부	점자블 록유무	집중조명 시설유무	관리기관 명	관리기관전 화번호	데이터기 준일자	acc_cnt	sch_cnt	old_cnt	ch_cnt	
0	충청 남도	서산 시	안건로	NaN	충청남도 서산시 갈산 동 163-3	s1	1	N	N	36.795415	...	N	N	N	충청남도 서산시청	041-660-3173	2022-11-16	1	0.0	0	0
1	충청 남도	서산 시	강경별 로	NaN	충청남도 서산시 부석 면 위평리 665-3	s2	1	N	N	36.720521	...	N	N	N	충청남도 서산시청	041-660-3173	2022-11-16	1	2.0	0	0
2	충청 남도	서산 시	고운로	NaN	충청남도 서산시 동문 동 53-3	s3	1	N	N	36.784331	...	Y	Y	N	충청남도 서산시청	041-660-3173	2022-11-16	28	4.0	0	0

[사진 6] sch\_cnt, ch\_cnt, old\_cnt 추가 데이터 예시

### d. 횡단보도별 인근 버스정류장 개수 데이터 추가

이외에도 시내의 교통사항 등을 반영하기 위하여 버스정류장의 개수를 추가하여 횡단보도 150m이내에 있는 버스정류장의 개수를 bus\_cnt로서 추가하였다.

	시도 명	시군 구명	도로 명	소재지도 로명주소	소재지지번주소	횡단보도관리 번호	횡단보 도종류	자전거횡단 도겸용여부	고원식 적용여 부	위도	...	점자블 록유무	집중조명 시설유무	관리기관명	관리기관 전화번호	데이터기 준일자	acc_cnt	sch_cnt	child_cnt	old_cnt	bus_cnt
0	충청 남도	천안 시	봉명 8길	NaN	충청남도 천안시 동남 구 봉명동 41-1	2018101016	1	NaN	N	36.806320	...	N	NaN	충청남도 천 안시 교통정 책과	041-521- 5893	2022-06- 14	26	10	11	1	2
1	충청 남도	천안 시	봉명 로	NaN	충청남도 천안시 동남 구 봉명동 59-20	2018101015	1	NaN	N	36.807002	...	Y	NaN	충청남도 천 안시 교통정 책과	041-521- 5893	2022-06- 14	27	11	11	1	2
2	충청 남도	천안 시	봉명 로	NaN	충청남도 천안시 동남 구 봉명동 59-20	2018101014	1	NaN	N	36.806933	...	Y	NaN	충청남도 천 안시 교통정 책과	041-521- 5893	2022-06- 14	29	11	11	1	2
3	충청 남도	천안 시	봉명 4길	NaN	충청남도 천안시 동남 구 봉명동 58-40	2018101013	1	NaN	Y	36.806384	...	Y	NaN	충청남도 천 안시 교통정 책과	041-521- 5893	2022-06- 14	26	10	11	1	2

[사진 7] bus\_cnt 추가 데이터 예시

### e. 횡단보도별 인근 아파트 세대수, 단지수 데이터 추가

교통사항 외에도 인구수에 따른 교통혼잡도 및 차량 통행여부를 반영하고자 대전, 세종, 충남의 아파트 단지 데이터를 활용하여 도로명, 지번주소를 이용하여 횡단보도 데이터와 지번, 도로명 주소가 일치하는 아파트 세대수와 단지수를 도로명은 apar\_rd\_cnt, apar\_rd\_fam으로 지번주소의 경우는 apar\_ad\_cnt, apar\_ad\_fam으로 추가하였다.

시도명	시군구명	도로명	소재지도로명주소	소재지지번주소	횡단보도관리번호	횡단보도종류	자전거횡단보도경유여부	고양식객용여부	위도	...	acc_cnt	sch_cnt	child_cnt	old_cnt	bus_cnt	apar_rd_cnt	apar_rd_fam	apar_ad_cnt	apar_ad_fam
0	충청남도 천안시	봉명8길	NaN	충청남도천안시동남구봉명동41-1	2018101016	1	NaN	N	36.806320	...	26	10	11	1	2	0	0	0	0
1	충청남도 천안시	봉명로	NaN	충청남도천안시동남구봉명동59-20	2018101015	1	NaN	N	36.807002	...	27	11	11	1	2	0	0	0	0
2	충청남도 천안시	봉명로	NaN	충청남도천안시동남구봉명동59-20	2018101014	1	NaN	N	36.806933	...	29	11	11	1	2	0	0	0	0
3	충청남도 천안시	봉명4길	NaN	충청남도천안시동남구봉명동58-40	2018101013	1	NaN	Y	36.806384	...	26	10	11	1	2	0	0	0	0
4	충청남도 천안시	봉명4길	NaN	충청남도천안시동남구봉명동517	2018101012	1	NaN	Y	36.806283	...	20	10	12	1	2	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
14581	세종특별자치시	갈매로	NaN	세종특별자치시어진동654	S100044	1	NaN	N	36.496807	...	4	5	3	0	3	2	1258	9	4282
14582	세종특별자치시	마음로	NaN	세종특별자치시고운동산95	S100045	1	NaN	N	36.521554	...	4	5	1	0	0	7	4156	21	12094
14583	세종특별자치시	마음로	NaN	세종특별자치시고운동산95	S100046	1	NaN	N	36.521554	...	4	5	1	0	0	7	4156	21	12094
14584	세종특별자치시	마음로	NaN	세종특별자치시고운동산95	S100047	1	NaN	N	36.521554	...	4	5	1	0	0	7	4156	21	12094
14585	세종특별자치시	마음로	NaN	세종특별자치시고운동산95	S100048	1	NaN	N	36.521554	...	4	5	1	0	0	7	4156	21	12094

14586 rows x 36 columns

[사진 8] 세대수, 단지수 추가 데이터 예시

이후 도로명과 지번주소 중 더 높은 세대수와 단지수를 통합하여 apar\_fam\_both, apar\_count\_both열로 저장하였다.

시도명	시군구명	도로명	소재지도로명주소	소재지지번주소	횡단보도관리번호	횡단보도종류	자전거횡단보도경유여부	고양식객용여부	위도	...	child_cnt	old_cnt	bus_cnt	apar_rd_cnt	apar_rd_fam	apar_ad_cnt	apar_ad_fam	cctv_cnt	apar_count_both	apar_fam_both
0	충청남도 천안시	봉명8길	NaN	충청남도천안시동남구봉명동41-1	2018101016	1	NaN	0	3.632476	...	2.484907	0.693147	1.098612	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000
1	충청남도 천안시	봉명로	NaN	충청남도천안시동남구봉명동59-20	2018101015	1	NaN	0	3.632494	...	2.484907	0.693147	1.098612	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000
2	충청남도 천안시	봉명로	NaN	충청남도천안시동남구봉명동59-20	2018101014	1	NaN	0	3.632493	...	2.484907	0.693147	1.098612	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000
3	충청남도 천안시	봉명4길	NaN	충청남도천안시동남구봉명동58-40	2018101013	1	NaN	1	3.632478	...	2.484907	0.693147	1.098612	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000
4	충청남도 천안시	봉명4길	NaN	충청남도천안시동남구봉명동517	2018101012	1	NaN	1	3.632475	...	2.564949	0.693147	1.098612	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
14581	세종특별자치시	갈매로	NaN	세종특별자치시어진동654	S100044	1	NaN	0	3.624256	...	1.386294	0.000000	1.386294	1.098612	7.138073	2.302585	8.362409	0.0	2.302585	8.362409
14582	세종특별자치시	마음로	NaN	세종특별자치시고운동산95	S100045	1	NaN	0	3.624916	...	0.693147	0.000000	0.000000	2.079442	8.332549	3.091042	9.400547	0.0	3.091042	9.400547
14583	세종특별자치시	마음로	NaN	세종특별자치시고운동산95	S100046	1	NaN	0	3.624916	...	0.693147	0.000000	0.000000	2.079442	8.332549	3.091042	9.400547	0.0	3.091042	9.400547
14584	세종특별자치시	마음로	NaN	세종특별자치시고운동산95	S100047	1	NaN	0	3.624916	...	0.693147	0.000000	0.000000	2.079442	8.332549	3.091042	9.400547	0.0	3.091042	9.400547
14585	세종특별자치시	마음로	NaN	세종특별자치시고운동산95	S100048	1	NaN	0	3.624916	...	0.693147	0.000000	0.000000	2.079442	8.332549	3.091042	9.400547	0.0	3.091042	9.400547

14586 rows x 38 columns

[사진 9] 세대수, 단지수 병합 데이터 예시

## f. 횡단보도별 인근 무인단속카메라 개수 데이터 추가

마지막으로 횡단보도 인근에 설치되어 있는 과속단속 카메라를 통한 영향이 있는지를 확인하기 위하여 공공데이터포털에서 대전, 충남, 세종의 무인단속카메라 데이터를 확보하여 횡단보도에 100m이내에 있는 무인단속카메라의 개수를 cctv\_cnt로 추가하였다.

	시도명	시군구명	도로명	소재지도로명주소	소재지지번주소	횡단보도관리번호	횡단보도종류	자전거횡단도겸용여부	고원식적용여부	위도	...	acc_cnt	sch_cnt	chi_id_cnt	old_cnt	bus_cnt	apar_fld_cnt	apar_fld_fam	apar_ad_cnt	apar_ad_fam	cctv_cnt
0	충청남도	천안시	봉명8길	NaN	충청남도천안시동남구봉명동41-1	2018101016	1	NaN	N	36.806320	..	26	10	11	1	2	0	0	0	0	0.0
1	충청남도	천안시	봉명로	NaN	충청남도천안시동남구봉명동59-20	2018101015	1	NaN	N	36.807002	..	27	11	11	1	2	0	0	0	0	0.0
2	충청남도	천안시	봉명로	NaN	충청남도천안시동남구봉명동59-20	2018101014	1	NaN	N	36.806933	..	29	11	11	1	2	0	0	0	0	0.0
3	충청남도	천안시	봉명4길	NaN	충청남도천안시동남구봉명동50-40	2018101013	1	NaN	Y	36.806384	..	26	10	11	1	2	0	0	0	0	0.0
4	충청남도	천안시	봉명4길	NaN	충청남도천안시동남구봉명동517	2018101012	1	NaN	Y	36.806283	..	20	10	12	1	2	0	0	0	0	0.0
...	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
14581	세종특별자치시	얼음	갈매로	NaN	세종특별자치시어진동654	SJ00044	1	NaN	N	36.496807	..	4	5	3	0	3	2	1258	9	4282	0.0

[사진 10] cctv\_cnt 데이터 예시

## g. 최종 데이터

이후 f까지의 과정을 거친 데이터들 중 군집화, 분류 모델에 사용될 변수들을 필터링하였으며, 목록은 아래 표와 같다.

열 이름	비고	열 이름	비고
시도명		교통섬유무	교통섬유무 Y/N
시군구명		차로수	
도로명	횡단보도 위치 도로 명칭	보도턱낮춤여부	보도턱낮춤여부 Y/N
소재지도로명주소		점자블럭유무	점자블럭유무 Y/N
소재지지번주소		acc_cnt	횡단보도 사고건수
횡단보도관리번호		sch_cnt	횡단보도 1km 이내 학교수
횡단보도종류	일반, 대각선, 이중 각각 1,2,3	bus_cnt	횡단보도 1km 이내 버스정류장수
고원식적용여부	고원식 여부 Y/N	cctv_cnt	횡단보도 100m 이내 cctv수
자전거횡단도 겸용여부	자전거횡단도 겸용여부 Y/N	old_cnt	횡단보도 1km 이내 노인시설수
위도	횡단보도 위도	apar_fam_both	횡단보도 인근 아파트 세대수

경도	횡단보도 경도	apar_count_both	횡단보도 인근 아파트 단지수
횡단보도연장	횡단보도 연장 길이 정보(m)		
횡단보도폭			
보행자신호등유무	보행자신호등유무 Y/N		

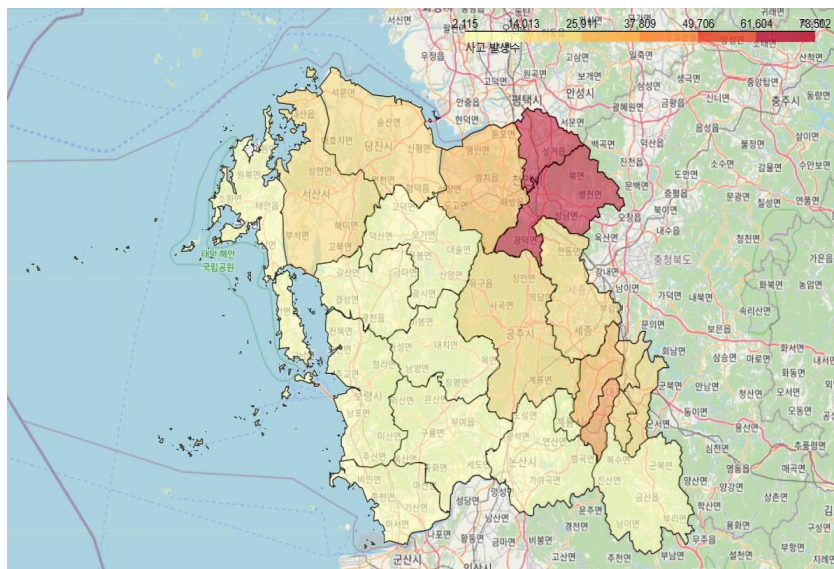
[표 1] 최종 데이터

## 5. 분석/시각화 결과 상세내용(EDA)

### a. map

EDA에 앞서, 더욱 자세히 확인할 수 있도록, 보고서 외 plot, map 등이 포함된 전체 코드를 참고용으로 함께 제공하였다. 횡단보도 관련 사고를 분석하기 위해 전국횡단보도표준데이터를 수집하였지만, 3.a Data에서 설명한 것과 같이 충청남도의 경우 일부 지역(천안시, 서산시, 보령시, 태안군, 서천군, 예산군)만 수집이 가능하였다. 본 연구에서 피쳐로 고려하고자 하는 아동과 노인인구, 학교 시설물 뿐만 아니라 사고 빈도를 map 형태로 나타낸 것은 다음과 같다.

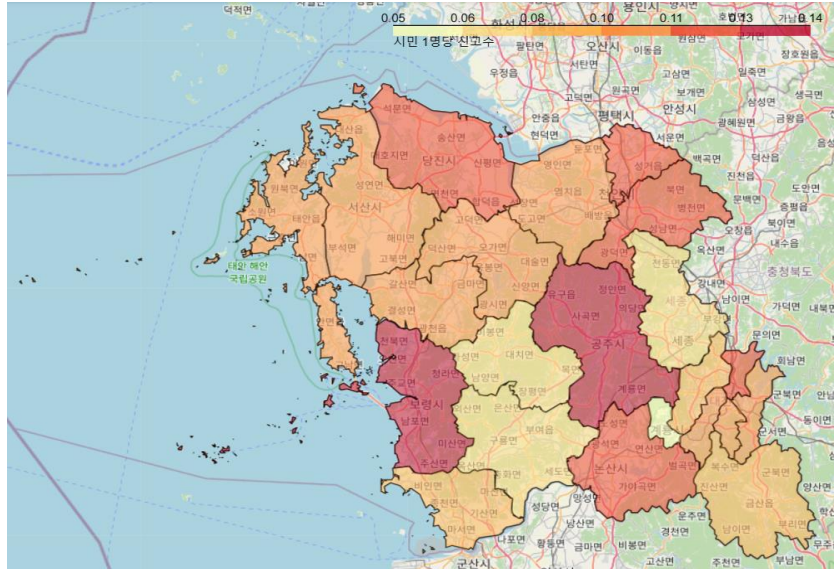
#### i. 교통사고 발생 건수



[사진 11] 교통사고 발생 건수 map

충청남도 천안시가 다른 지역들에 비해 교통사고 발생 건수가 많고, 그 외 대전광역시, 세종특별자치시, 충청남도 아산시, 서산시, 당진시, 공주시가 비교적 교통사고 발생 건수가 많음을 알 수 있다.

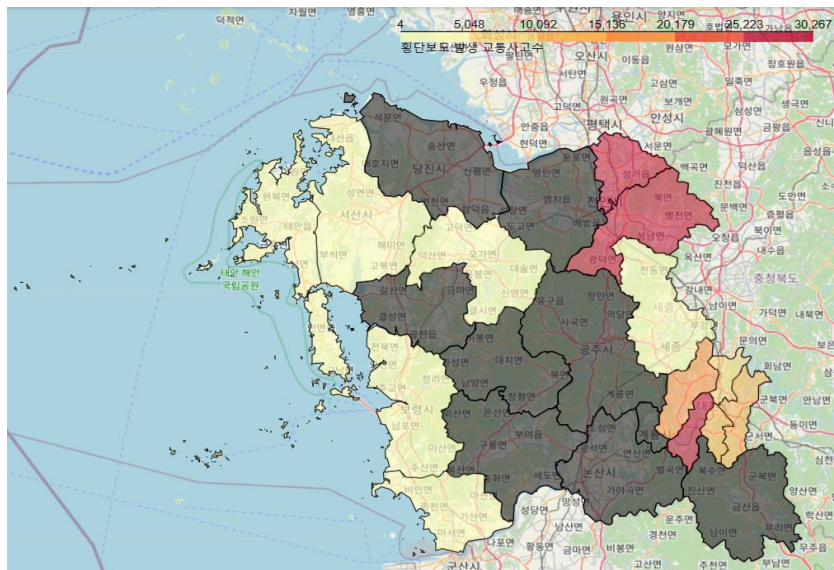
## ii. 인구 1명 당 교통사고 발생 건수



[사진 12] 인구 1명 당 교통사고 발생 건수 map

충청남도 보령시, 공주시가 1인 당 교통사고 발생 건수가 높은 편임을 알 수 있다.

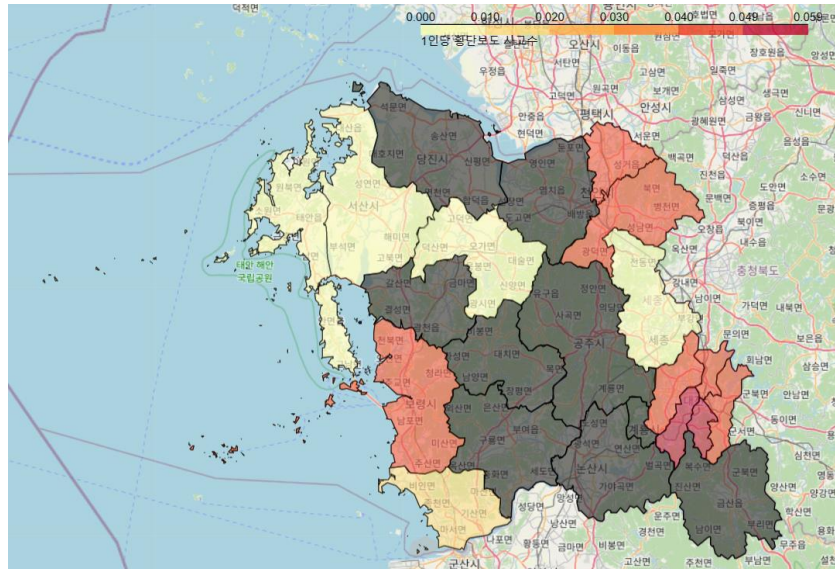
## iii. 횡단보도 근방 교통사고 건수



[사진 13] 횡단보도 근방 교통사고 발생 건수 map

횡단보도표준데이터를 구할 수 없었던 충청남도 일부 지역은 검은색으로 처리되었다. 충청남도 지역 중 교통사고 발생 건수 혹은 인구 1명 당 교통사고 발생 건수가 높은 지역들(천안시, 보령시, 서산시, 서천군 등)이 다수 포함되었고, 이 map을 통해 충청남도 천안시, 대전광역시 서구가 횡단보도 근방 사고 건수가 많음을 알 수 있다.

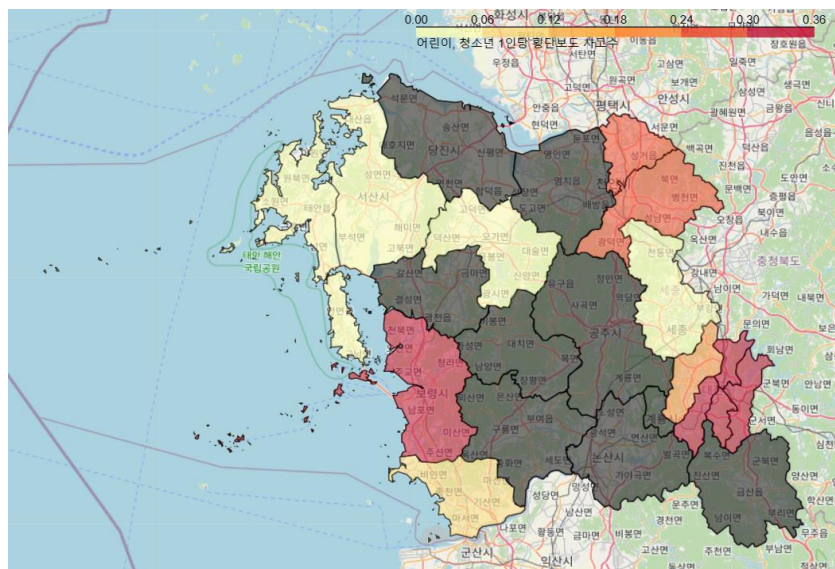
#### iv. 인구 1명 당 횡단보도 근방 교통사고 발생 건수



[사진 14] 인구 1명 당 횡단보도 근방 교통사고 발생 건수 map

인구 1명 당 횡단보도 근방 교통사고 발생 건수의 경우, 대전광역시, 충청남도 천안시, 보령시가 비교적 높음을 알 수 있다.

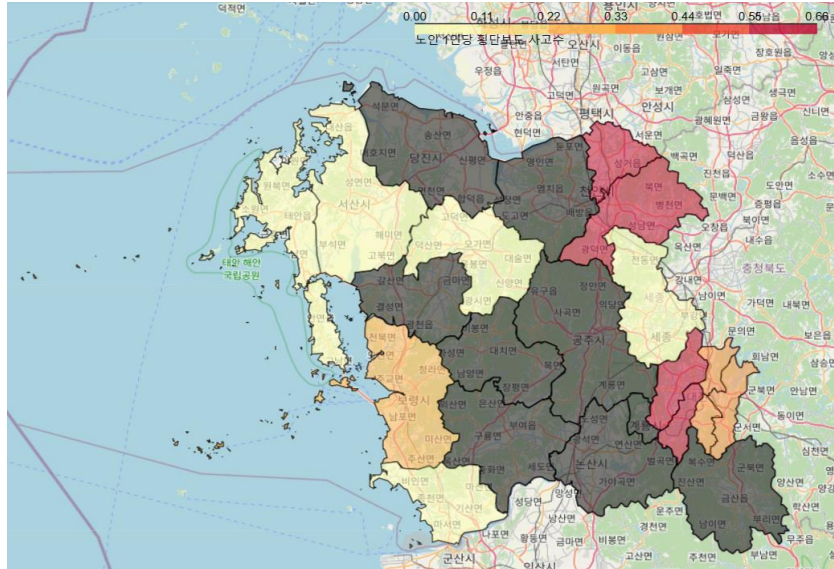
#### v. 아동·청소년 인구 1명 당 횡단보도 근방 교통사고 발생 건수



[사진 15] 아동·청소년 인구 1명 당 횡단보도 근방 교통사고 발생 건수 map

‘스마트 횡단보도’ 도입에 있어 아동·청소년을 고려하여 EDA를 진행한 결과, 대전광역시의 대다수 구와 충청남도 보령시의 경우 ‘인구 1명 당 횡단보도 근방 교통사고 발생 건수’보다 더욱 높아진 것을 알 수 있다.

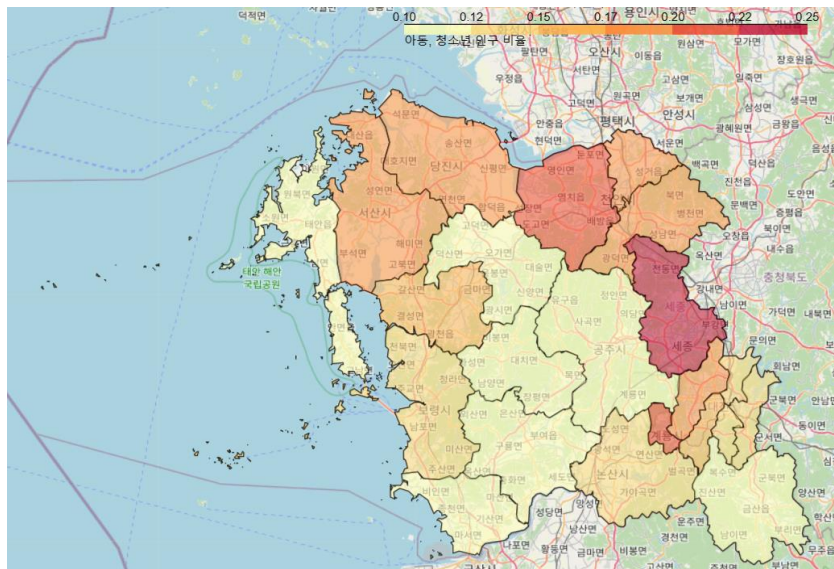
#### vi. 노인 인구 1명 당 횡단보도 근방 교통사고 발생 건수



[사진 16] 노인 인구 1명 당 횡단보도 근방 교통사고 발생 건수 map

‘스마트 횡단보도’ 도입에 있어 노인을 고려하여 EDA를 진행한 결과, 충청남도 천안시, 대전광역시 유성구의 경우 ‘인구 1명 당 횡단보도 근방 교통사고 발생 건수’보다 더욱 높아진 것을 알 수 있다.

#### vii. 아동·청소년 인구 비율

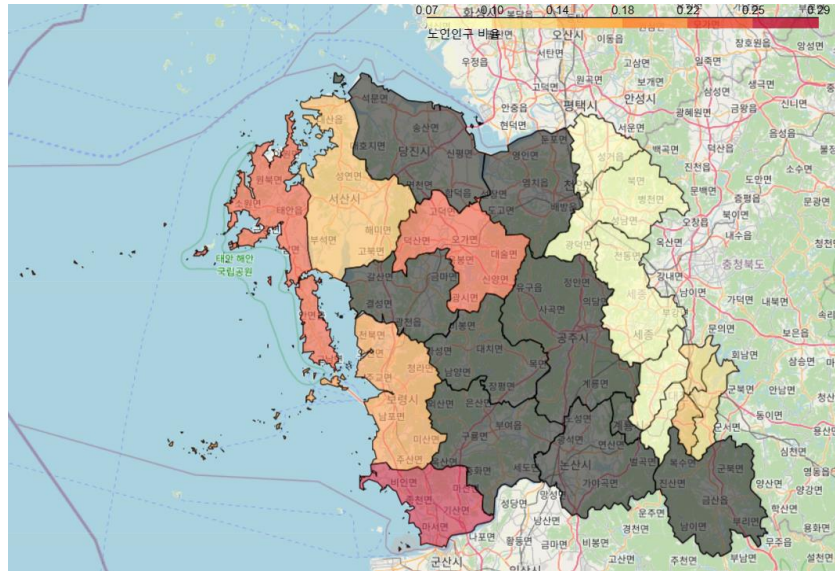


[사진 17] 아동·청소년 인구 비율 map

세종특별자치시가 아동·청소년 인구 비율이 매우 높은 것을 알 수 있다.

횡단보도표준데이터를 구할 수 있었던 충청남도 천안시, 서산시, 보령시의 경우도 타 충청남도 지역 보다 비교적 아동·청소년 인구 비율이 높게 나타났다.

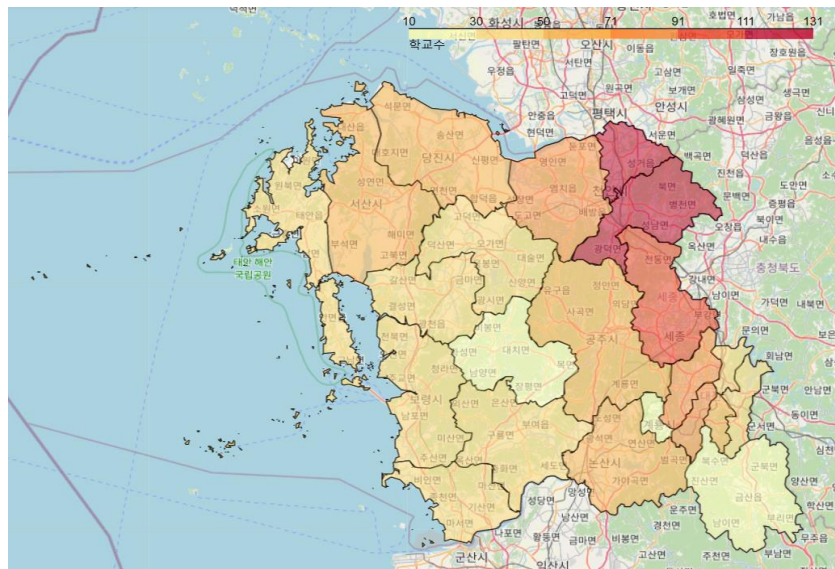
#### viii. 노인 인구 비율



[사진 18] 노인 인구 비율 map

충청남도 천안시를 제외한 지역들이 노인 인구 비율이 높은 것을 알 수 있다. 즉, 횡단보도표준데이터를 구할 수 있었던 충청남도 지역들 중 천안시를 제외하면 모두 노인 인구 비율이 높았다.

#### ix. 학교 수

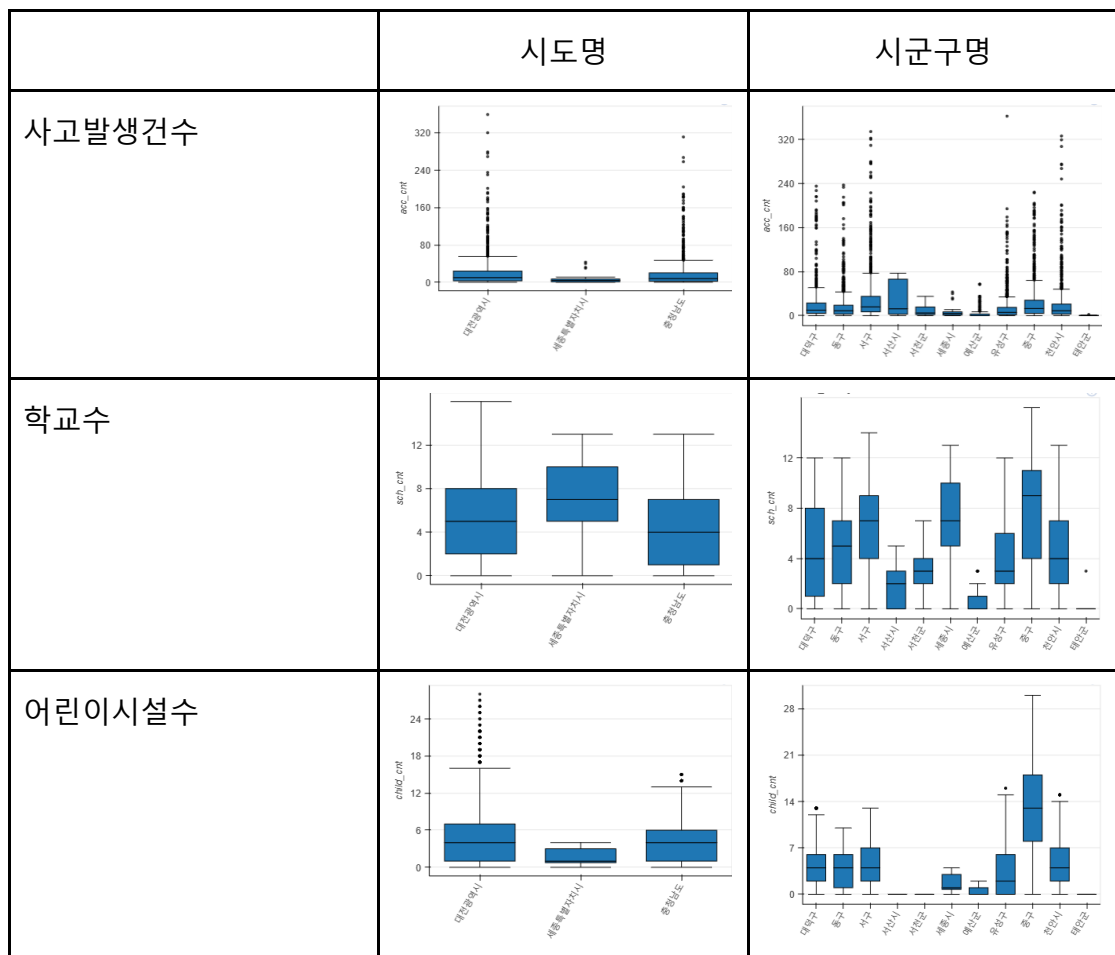


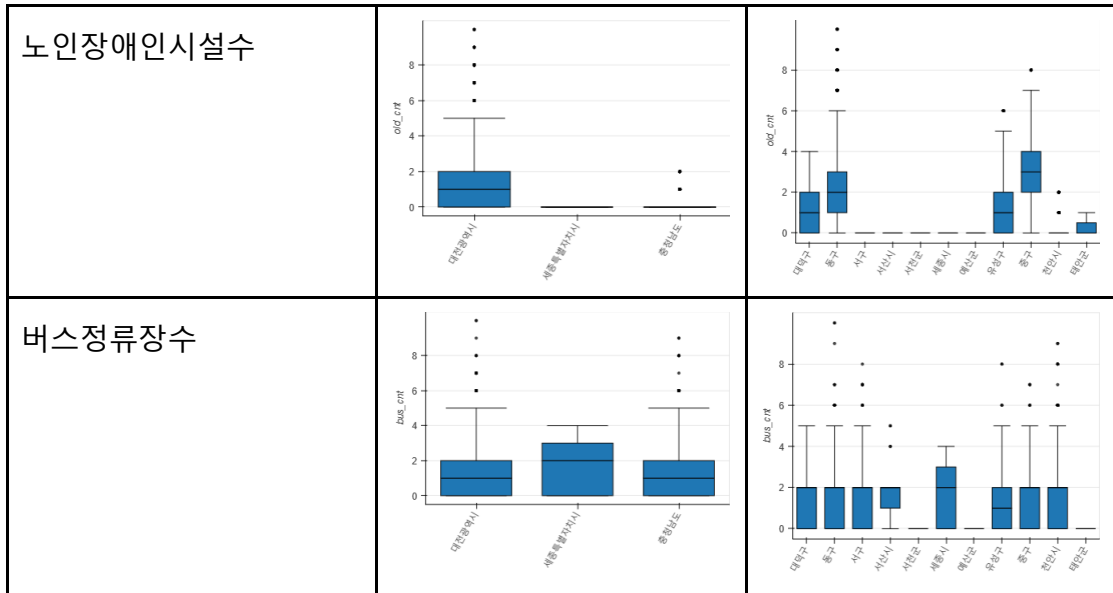
[사진 19] 학교 수 map

학교 수는 충청남도 천안시, 세종특별자치시가 높은 것을 확인할 수 있다. 마찬가지로 횡단보도표준데이터를 구할 수 있었던 충청남도 서산시도 충청남도 지역 내 학교 수가 많은 편에 속한다.

map을 통해 확인할 수 있었던 점은, 횡단보도를 구할 수 없는 충청남도 일부 지역들 중 충청남도 아산시, 당진시, 공주시를 제외하면 교통사고가 비교적 매우 드문 지역들이다. 또한, 사고 건수가 매우 높은 천안시의 횡단보도표준데이터가 존재함과 더불어 교통사고 건수와 아동·청소년 비율이 비교적 높은 서산시, 1인당 교통사고 건수와 노인 인구 비율이 매우 높은 보령시 등 유의미한 충청남도 지역들의 횡단보도표준데이터가 존재한다는 점을 고려하였을 때 충분히 모델링 및 '스마트 횡단보도' 최적 입지 선정이 가능하다고 판단하였다.

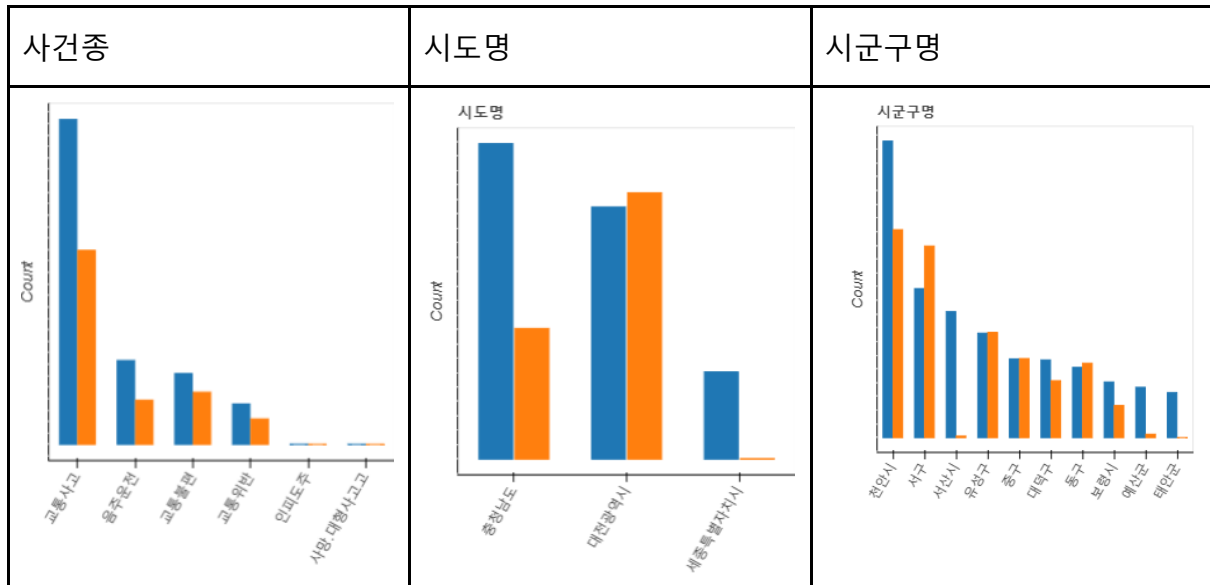
## b. plot





[표 2] EDA

횡단보도에서 발생한 사고를 제외한 교통사고 관련 신고 데이터(파란색)과 횡단보도에서 발생한 사고 데이터의 수(주황색)는 다음과 같다. 여기서 교통사고 관련 신고 데이터란 교통사고, 음주운전, 교통불편, 교통위반, 인피도주, 사망 및 대형사고의 신고 데이터를 의미한다.



[표 3] EDA

사건종별 횡단보도에서 발생하는 교통사고 관련 신고건수와 횡단보도를 제외한 곳에서 발생하는 신고건수의 수를 비교해보았을 때, 횡단보도에서 발생한 사고 신고건수가 결코 적지 않음을 알 수 있다. 상당수의 교통관련 사고가 횡단보도에서 이루어진다는 판단 하에, 이를 줄이고자 '스마트 횡단보도'라는 개선방안을 도출하였다. 하지만 예산 등 현실적인 문제로 인하여 모든 횡단보도에 스마트

횡단보도를 도입하는 것은 사실 상 불가능하다. 따라서 최소한의 스마트 횡단보도만으로도 최대한 많은 교통관련 사고를 줄이는데 기여하고자 최적입지를 선정하고자 한다. 최적입지 선정을 위해 사고건수를 비롯하여 취약계층인 어린이,노인,장애인을 위해 인근의 관련 시설수, 학교수, 버스정류장수를 추가로 고려하였다.

궁극적인 목적이 스마트 횡단보도의 설치를 위한 최적입지를 선정하는 것이므로, 횡단보도에서 발생한 사고만을 고려하여 분석을 진행하였다. 사고다발지역을 선정하는 과정을 통해 결론적으로 선정된 최적입지는 전체 교통관련 사고 신고건수가 아닌, 횡단보도에서 발생한 사고의 신고건수이다. 예를 들어 충청남도 서산시와 대전광역시 동구를 비교해보았을 때, 총 발생한 교통관련 신고건수는 서산시와 동구가 비슷하다. 하지만 횡단보도에서 발생한 교통관련 신고건수는 동구가 월등히 높다. 이러한 경우, 동구가 보행자 교통 사고에 취약한 지역임을 알 수 있다.

## 6. 모델링 & 결과 해석

### a. 사고다발지역 예측 모델

앞서 데이터 전처리 과정을 통해 각 횡단보도별 신고 및 사고 건수(acc\_cnt)를 추출할 수 있었다. 해당 열을 10 이상이면 '1', 10미만이면 '0'으로 변환하여 10회 이상 신고가 발생한 지역을 '사고다발지역'으로 구분하였다. 해당 데이터의 분포는 아래의 표와 같으며 사고다발지역으로 분류된 데이터는 전체의 약 49%이다.

사고다발지역	데이터 수
0	7462
1	7124

[표 4] 데이터의 레이블 분포

따라서 총 14,586개의 횡단보도 중 사고다발지역인 횡단보도를 분류하는 예측 모델링을 진행하였다. 선정된 모델은 여러 모델 중 랜덤포레스트가 가장 높았으며

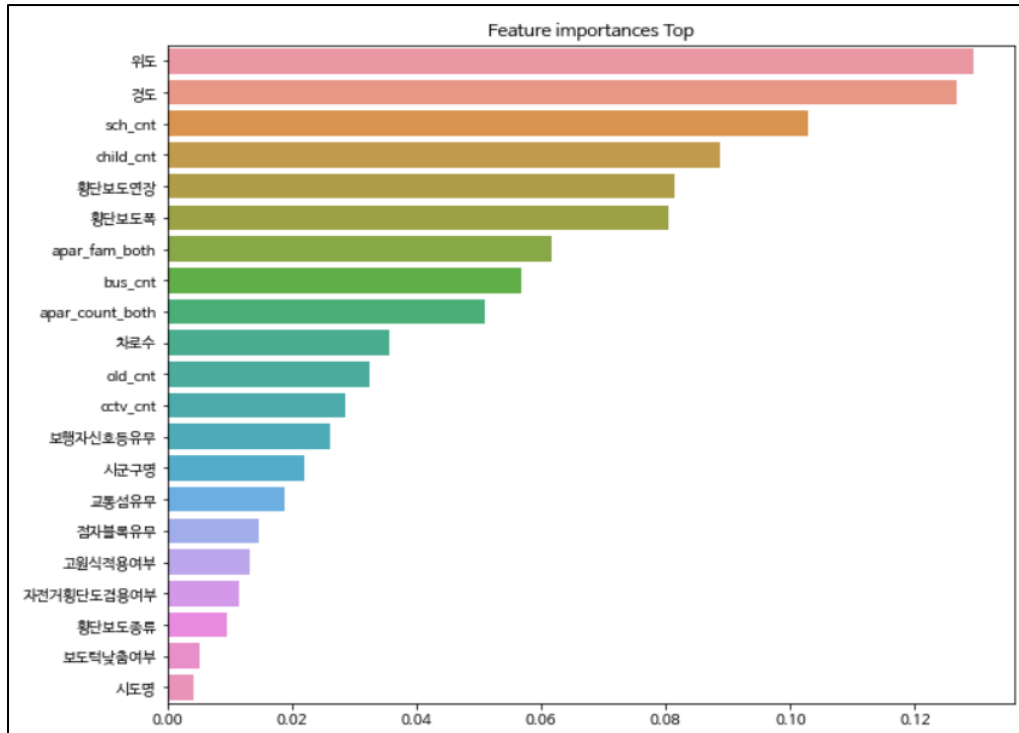
해당 모델에 대해 추가적으로 하이퍼 파라미터를 조정하는 과정을 거쳐 최적값을 도출하였다. 결과는 아래의 표와 같다.

평가지표	Randomforest	GridsearchCV 최적값
accuracy	0.83	0.838
recall	0.851	<b>0.861</b>
precision	0.815	0.82
F1 score	0.832	0.84

[표 5] 모델링 최종 결과값

분류 결과 실제 사고다발지역인 횡단보도 중 모델이 사고다발지역이라고 판단한 것(recall)의 비율은 86%임을 확인할 수 있다. 따라서 해당 모델을 통해 다른 지역의 사고다발지역을 판단하거나 새로운 횡단보도를 설치하는데 의사결정의 근거가 될 수 있을 것이다.

다음으로 해당 모델에서 사고다발지역 분류에 있어서 중요한 피처를 순서대로 나타내보았다. 결과를 살펴보면 횡단보도의 위치를 나타내는 '위도', '경도' 값이 가장 중요하며 다음으로 주위 시설수와 횡단보도의 특징을 나타내는 열들에 의해 분류된다는 것을 알 수 있다. 하지만 횡단보도의 위도, 경도를 통해 사고다발지역을 예측하는 것은 횡단보도의 특징과 지역의 특성을 반영하는 것이 아닌 위치의 영향이 크다고 판단하였고, 추가적인 모델링이 필요하다고 판단하였다.



[그래프 2] 최적 예측 모델에 대한 피쳐 중요도

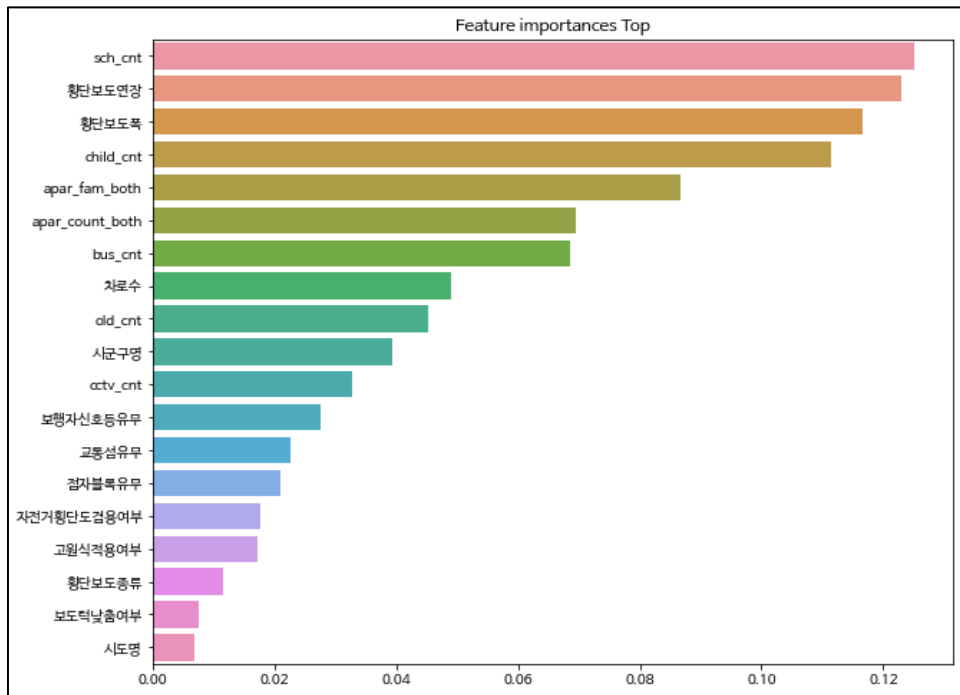
## b. 피쳐 중요도 추출

위에서 언급했듯이 횡단보도의 특성과 주변 환경 변수만을 고려하여 피쳐 중요도를 추출하고자한다. 따라서 횡단보도의 위치를 특정할 수 있는 '위도', '경도' 변수는 제거하고 다시 모델링을 진행하였다. 모델링 결과는 아래의 표와 같다.

평가지표	Randomforest	GridsearchCV 최적값
accuracy	0.809	0.812
recall	0.829	<b>0.836</b>
precision	0.794	0.795
F1 score	0.811	0.815

[표 6] 위도, 경도 변수 제거 결과값

예측 성능은 위치 정보를 포함한 a 단계의 모델에 비해 떨어졌지만 그럼에도 84%의 높은 성능을 가짐을 확인하였다. 다음으로 모델의 예측에 있어서 중요한 피쳐는 [그래프 3]과 같다.

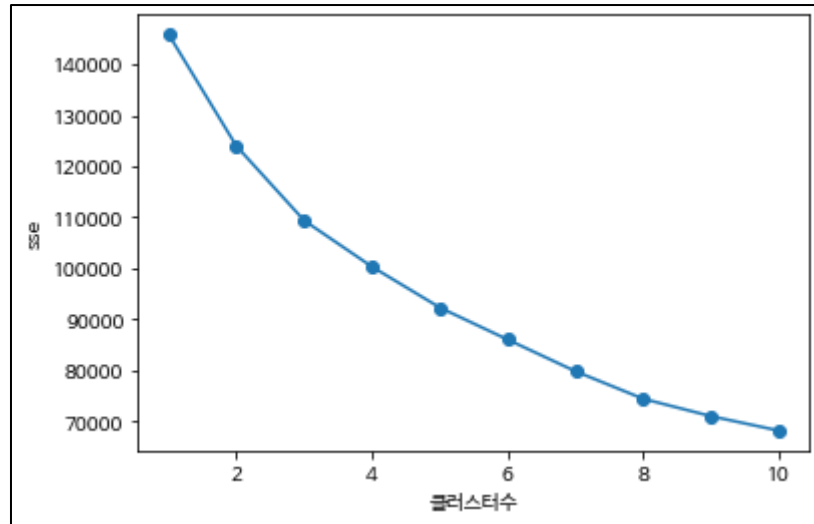


[그래프 3] 위도, 경도를 제외한 피처 중요도

중요도를 시각화한 결과 주변 시설에 관한 피처 중에서는 '학교 수', '어린이 시설 수', '아파트 세대 수', '버스 정류장 수', '노인 시설 수', 'cctv 수' 순으로 중요한 영향을 미친 것을 확인할 수 있었다. 또한 횡단보도의 특성에 관한 피처 중에서는 '횡단보도연장', '횡단보도폭', '차로수' 순으로 사고다발 지역을 결정 짓는 중요한 역할을 했음을 확인할 수 있다. 횡단보도의 특성을 나타내는 '보행자신호등유무', '교통섬유무'와 같은 변수도 존재하지만 중요도가 상대적으로 떨어지기 때문에 클러스터링을 위한 변수로는 '차로수'까지만 고려하기로 결정하였다.

### c. 클러스터링

예측 모델링과 피처 중요도를 확인하는 과정에서 사고다발지역과 관련된 중요한 피처들을 선별하였다. 하지만 우리의 목적은 단순히 사고다발지역을 선정하는 것이 아닌 '어린이'와 '노약자' 보행자에게 위험한 '보행자 교통 취약 지역'을 선정하는 것이다. 따라서 주변 환경을 고려한 피처들과 교통사고에 영향을 주는 피처들을 사용하여 클러스터링함으로써 우리의 목적에 맞는 지역을 군집화하고 각 군집의 특징을 살펴보고자한다. 클러스터링의 방법으로는 k-means 방법을 사용하였으며, 먼저 최적 군집 수를 찾기 위해 elbow method를 사용하여 시각화하였다.



[그래프 4] elbow method를 사용하여 나타낸 각 군집 개수에 따른 변동 정도

[그래프 4]를 통해 군집이 3개일 때와 8개일 때 변동성이 소폭 감소하였음을 확인하였다. 따라서 먼저 3개의 군집인 경우 각 피처의 평균 특징을 살펴보았다.

	신고수	학교수	어린이시설수	노인시설수	아파트세대수	cctv수	버스정류장수	횡단보도폭	횡단보도연장	차로수
predict										
0	49.20	5.47	5.03	0.93	6335.21	0.36	1.98	8.68	17.76	5.78
1	8.33	2.46	2.30	0.61	1847.92	0.08	0.84	4.85	8.71	2.11
2	22.79	7.83	7.73	1.64	5965.78	0.37	1.58	5.02	8.23	2.05

[데이터프레임 1] 각 군집에 대한 평균 특성

위와 같이 3개의 군집일 경우 사고수, 학교수, 어린이 시설수, 노인시설수가 상대적으로 높은 2번 군집이 우리의 선정하고자 하는 입지와 가장 유사하다는 것을 알 수 있다. 하지만 조금 더 세분화하여 군집을 나누기 위해 8개로 군집을 설정하여 다시 결과를 확인하였다. 피처 중 값의 편차가 큰 데이터가 있기 때문에 평균과 중앙값을 활용하여 종합적으로 비교해보았다.

신고수	학교수	어린이시설수	노인시설수	아파트세대수	cctv수	버스정류장수	횡단보도폭	횡단보도연장	차로수	predict
20.221514	4.892431	4.709242	1.03332	4068.429659	0.232209	1.296517	5.536524	10.009487	2.688537	3.230289

[데이터프레임 2] 전체 데이터에 대한 각 열의 평균

	신고수	학교수	어린이시설수	노인시설수	아파트세대수	cctv수	버스정류장수	횡단보도폭	횡단보도연장	차로수
predict										
0	39.03	6.18	5.66	0.39	2739.65	0.36	1.89	17.75	6.87	4.78
1	6.33	1.70	1.46	0.64	1603.11	0.01	0.69	4.71	8.50	2.02
2	20.37	8.43	11.66	4.51	3791.49	0.24	1.53	5.16	10.19	2.33
3	23.82	8.09	4.14	0.09	19041.19	0.16	1.63	5.25	10.27	2.90
4	22.79	3.77	3.90	0.92	4204.52	0.15	1.65	6.96	21.43	5.90
5	16.50	6.52	6.35	0.63	3112.61	0.00	1.48	4.96	7.69	2.06
6	20.13	5.65	4.88	0.81	3141.18	1.26	1.50	5.17	8.64	2.36
7	163.54	6.65	7.19	1.49	5948.49	0.51	2.34	6.73	14.93	4.56

[데이터프레임 3] 각 군집에 대한 평균 특성

신고수	학교수	어린이시설수	노인시설수	아파트세대수	cctv수	버스정류장수	횡단보도폭	횡단보도연장	차로수	predict
9.0	4.0	4.0	0.0	2057.0	0.0	1.0	4.8	7.4	2.0	3.0

[데이터프레임 4] 전체 데이터에 대한 각 열의 중앙값

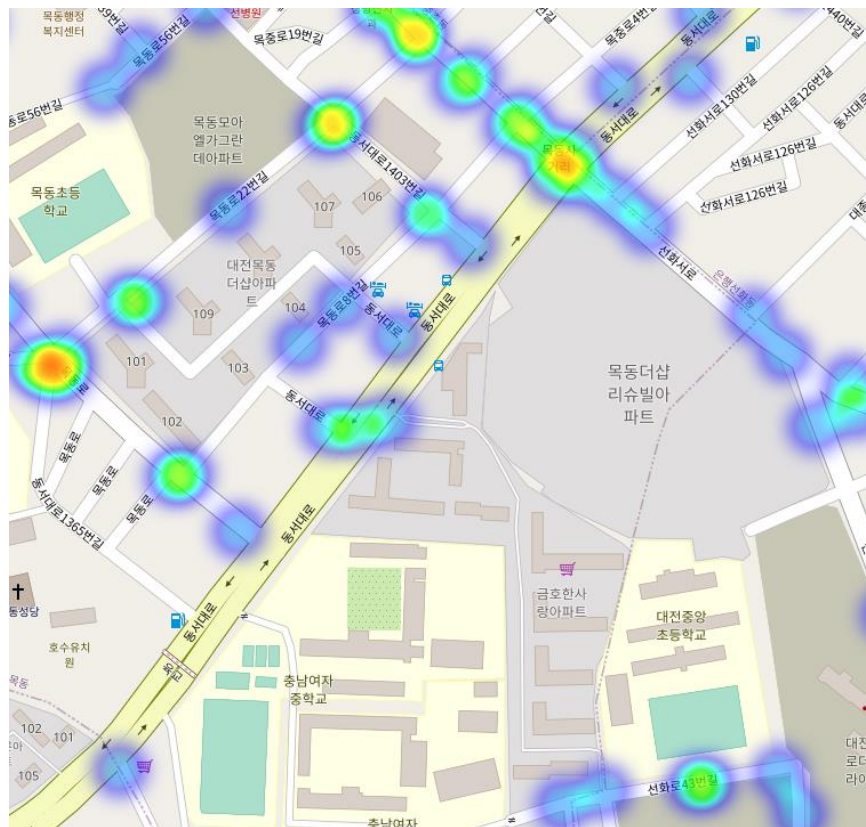
	신고수	학교수	어린이시설수	노인시설수	아파트세대수	cctv수	버스정류장수	횡단보도폭	횡단보도연장	차로수
predict										
0	21.0	7.0	5.0	0.0	1994.0	0.0	2.0	16.2	6.99	5.0
1	2.0	1.0	1.0	0.0	515.0	0.0	0.0	4.1	7.40	2.0
2	14.0	9.0	10.0	4.0	2616.0	0.0	2.0	4.6	8.45	2.0
3	17.0	8.0	4.0	0.0	22110.0	0.0	2.0	5.0	7.70	2.0
4	13.0	3.0	3.0	0.0	2634.0	0.0	2.0	6.6	20.40	6.0
5	10.0	7.0	6.0	0.0	2610.0	0.0	2.0	4.1	7.00	2.0
6	14.0	6.0	4.0	0.0	1804.5	1.0	2.0	4.6	7.08	2.0
7	153.0	6.0	6.0	1.0	3732.0	0.0	2.0	6.2	12.51	5.0

[데이터프레임 5] 각 군집에 대한 중앙값

먼저 군집의 결과를 통해 2번 군집을 우리의 목적에 가장 부합하는 군집이라고 판단하였다. 해당 군집과 같은 경우, 학교와 주위 시설의 수가 다른 군집에 비해 월등히 높은 것을 확인할 수 있다. 따라서 학교와 주거 공간이 적절히 섞인 지역으로 이로 인해 어린이와 노인을 위한 시설이 균형 있게 갖춰져 있으며, 일상생활에 밀접한 횡단보도임을 알 수 있다. 또한 횡단보도 폭과 차로 수가 크지 않음에도 주위 인프라로 인해 차가 많이 다니는 지역으로 예상되며 그로 인해 교통 관련 신고와 사고가 높은 것이라 추측된다. 따라서 해당 지역이 스마트 횡단보도를 도입했을 때 가장 효과가 높을 것으로 생각되어 해당 군집을 '어린이, 노인을 위한 보행자 취약 지역'으로 선정하였다.

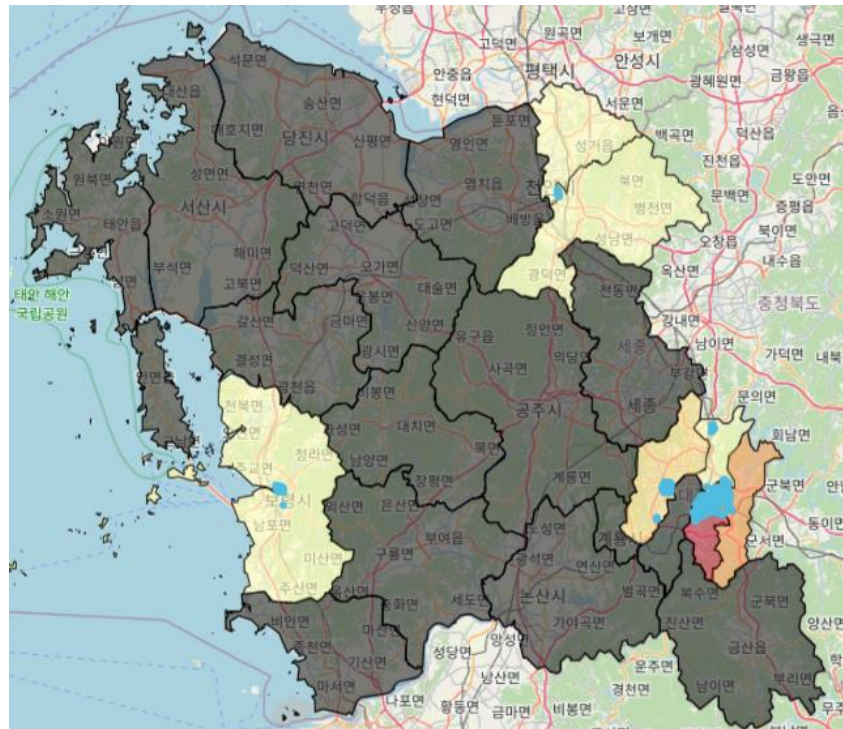


[사진 20] 대정광역시 중구 대흥동 630



[사진 21] 대전 중구 중촌동

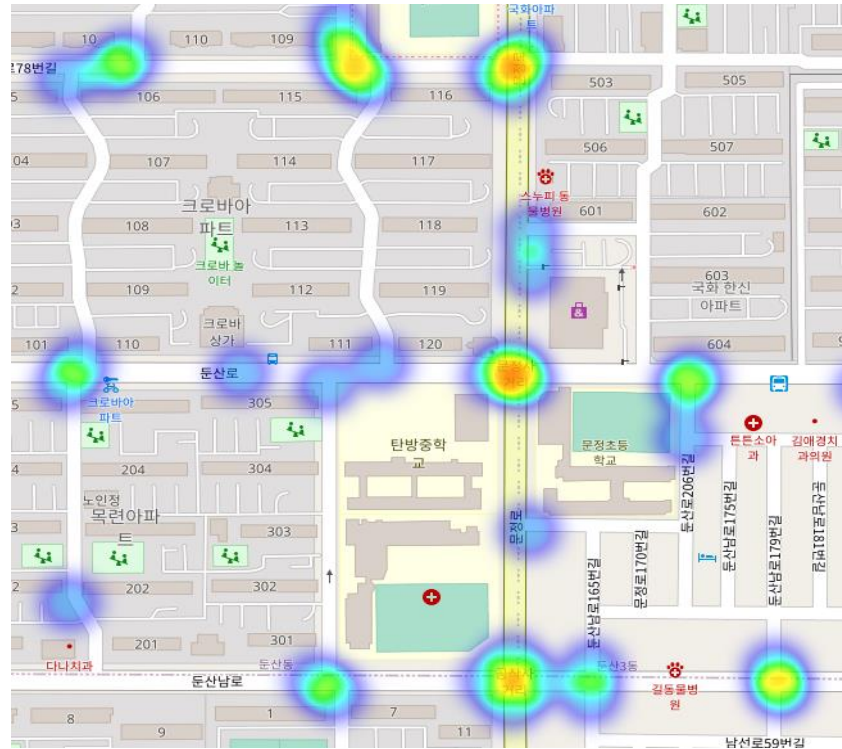
추가적으로 해당 클러스터 내 횡단보도를 각 지역에 해당하는 밀도로 표시하고 파란색 점으로 위치를 표현한 그림은 아래와 같다. 그림을 통해 대전광역시에도 많이 밀집되어 있다는 사실을 확인할 수 있다.



[사진 22] 대전 중구 중촌동

스마트 횡단보도를 위한 최적 군집은 2번으로 선택하였지만, 학교수가 높은 3번 군집과 신고수가 많은 7번 군집을 추가적으로 분석하여 차이를 비교해보았다.

3번 군집과 같은 경우 교통사고에 중요한 영향을 미치는 학교수가 2번 군집과 비슷하게 높지만 아파트 세대수의 값이 월등히 큰 것을 확인할 수 있다. 이는 아래의 히트맵과 같이 주거 밀집 지역에 위치한 학교 주변에서 발생하는 문제라고 판단된다. 하지만 이는 어린이와 노약자와 같은 보행자를 넓게 고려할 수 없기 때문에 스마트 횡단보도를 설치하는 것보다 아이들의 등하교 시간에 녹색어머니회 같은 활동을 하는 것이 더 효과적이라고 판단된다.



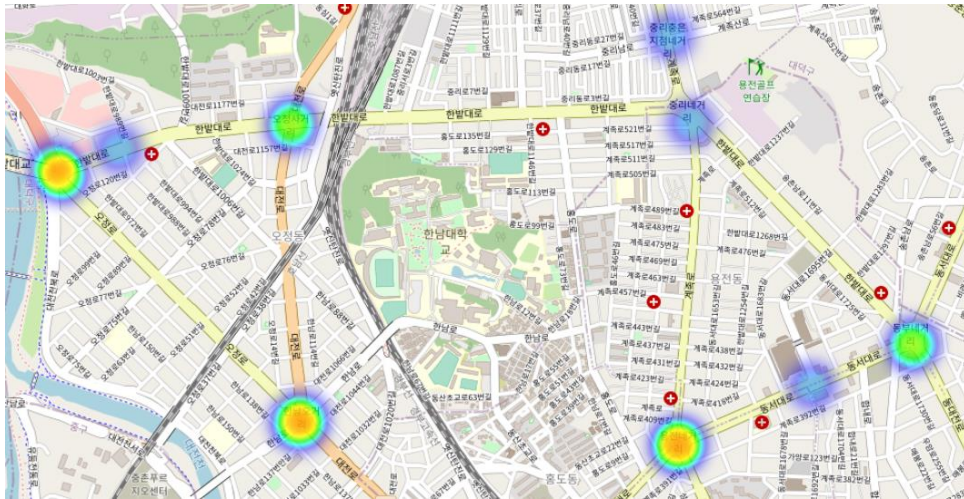
[사진 23] 대구 서구 문정로

해당 히트맵을 통해 교통이 혼잡하지 않은 주거 밀집지역 속 분포해있음을 알 수 있다.

다음으로 3번 군집과 같은 경우 횡단보도 폭과 차로수가 다른 군집에 크며 신고수가 월등히 높은 것을 통해 교통량이 많은 지역이라고 판단된다. 따라서 해당 군집 내 신고수가 가장 높은 횡단보도의 신고 종류를 살펴본 결과 총 362건 중 '교통불편(402)' 신고가 164건으로 확인되었다. (해당 신호등을 기준으로 반경 50m내 좌표에 위치한 신고 데이터) 따라서 해당 횡단보도, 즉 해당 군집은 보행자 교통사고 취약지역이라고 보기 어려우며, 운전자를 위한 교통 편의 개선이 필요하다고 판단된다.



[사진 24] 대전광역시 유성구 노은동 42-5



[사진 25] 대전광역시 유성구 노은동

해당 히트맵을 통해 주로 교통이 혼잡한 대형 교차로에 분포해있음을 알 수 있다.

#### d. 피쳐 중요도를 활용한 가중치 계산

위에서 선정한 군집 내 횡단보도 데이터를 활용하여 스마트 횡단보도 설치를 위한 우선 선정 지표를 제시하고자 한다. 선정 지표는 따라서 학교 및 어린이와 노약자를 위한 주변 시설에 대한 가중치와 신고수를 복합적으로 고려하는 식으로 설계하였다. 식을 구하는 과정은 다음과 같다.

<우선 선정 지표 계산>

1. b 단계 에서 구한 학교수, 어린이와 노인 시설 수의 피쳐 중요도를 추출한다

피쳐	중요도
학교 수	<b>0.125</b>
어린이 시설 수	<b>0.112</b>
노인 시설 수	<b>0.045</b>

[표 7] 각 피쳐에 대한 중요도

2. 횡단보도의 값과 해당 피쳐의 가중치를 곱하여 주위 시설에 대한 종합 고려 가중치를 계산한다. (각 피쳐에 대한 스케일링 진행 후 계산하였음.)

$$\text{종합 가중치} = \text{학교수} \times 0.125 + \text{어린이시설수} \times 0.112 + \text{노인시설수} \times 0.045$$

3. 주위 시설에 대한 종합 가중치와 신고수를 곱하여 최종 지표를 계산한다.

$$\text{우선 선정 지표} = \text{종합 가중치} \times \text{신고수}$$

4. 종합 가중치가 '음수'인 횡단보도는 제외한다.

종합 가중치가 '음수'인 경우는 전체적인 가중치에서 평균 미만임을 의미하기 때문에 선정 고려 대상에서 제외하였다. (군집 중 절반 정도 제외된다.)

해당 과정을 통해 구한 '우선 선정 지표'를 기준으로 정렬된 횡단보도 데이터 예시는 다음과 같다. 표의 결과를 살펴보면 단순히 신고건수가 낮더라도 주위 시설의 정도에 따라 높은 우선순위를 갖는 횡단보도를 확인할 수 있다.

소재지번주소	횡단보도관리번호	신고수	학교수	어린이시설수	노인시설수	종합가중치	우선선정지표
대전광역시중구오류동118-4	624	72	15	28	6	0.527726	37.996302
대전광역시중구중촌동418-1	7108	108	14	21	5	0.295476	31.911398
대전광역시중구중촌동1-144	7107	108	14	21	5	0.295476	31.911398
대전광역시중구오류동196-1	620	64	15	28	5	0.494878	31.672178
대전광역시중구중촌동156-35	7109	98	14	21	5	0.295476	28.956639

[데이터프레임 6] 대전광역시 중구

## 7. 결론

### a. 최적 입지 선정

본 연구는 5.c 단계의 클러스터링을 통해 총 1,394개의 어린이, 노약자를 위한 보행자 취약 지역을 1차로 선정하였다. 다음으로 d 단계에서 클러스터 속 모든 지역에 대해 통합적인 종합가중치를 구하고 음수 값을 가지는 횡단보도를 제외함으로써 603개를 2차로 선정할 수 있었다.

하지만 각 지역에 대해 좀 더 세부적으로 살펴보기 위해 지역을 나누어 우선 선정 지표를 적용하고 각 지역에 대한 '스마트 횡단보도' 최적 입지 1개씩을 예시로 제시하고자한다. 클러스터를 통해 1차로 선정된 횡단보도의 각 지역에 대한 분포는 다음과 같다.

지역	선정된 횡단보도 개수
대전광역시 중구	657
대전광역시 동구	364

대전광역시 유성구	184
대전광역시 대덕구	110
충청남도 보령시	64
충청남도 천안시	15

[표 8] 선정 군집 내 각 지역의 횡단보도 개수

### <대전광역시 중구>

먼저 대전광역시의 중구 지역을 추출하여 종합가중치와 우선선정지표를 구하였다. '우선선정지표'를 기준으로 정렬했을 때, 상위 2개의 횡단보도는 다음과 같다. (횡단보도 중 근처에 위치하여 신고수가 같게 집계된 횡단보도는 제외하였다. 각 지역에 대한 우선순위 결과는

[https://colab.research.google.com/drive/1MCNFzRjk7vxOK-zYAezdunixf5SJ\\_HW0G?usp=sharing](https://colab.research.google.com/drive/1MCNFzRjk7vxOK-zYAezdunixf5SJ_HW0G?usp=sharing) 에서 '군집화' 파트를 참고하면 된다.)

	소재지지번주소	횡단보도관리번호	위도	경도	신고수	학교수	어린이시설수	노인시설수	아파트세대수	차로수	종합가중치	우선선정지표
8	대전광역시중구오류동118-4	624	36.32356	127.404209	72	15	28	6	3193	1	0.527726	37.996302
411	대전광역시중구종로동418-1	7108	36.33474	127.413141	108	14	21	5	5236	3	0.295476	31.911398

[데이터프레임 7] 대전광역시 중구의 상위 2개 횡단보도

따라서 1순위로 나온 횡단보도를 거리뷰로 살펴본 사진은 아래와 같으며, 이미 '스마트 횡단보도'가 설치 되어있음을 확인하였다. 따라서 우선 선정 지표의 타당성을 입증할 수 있었다.

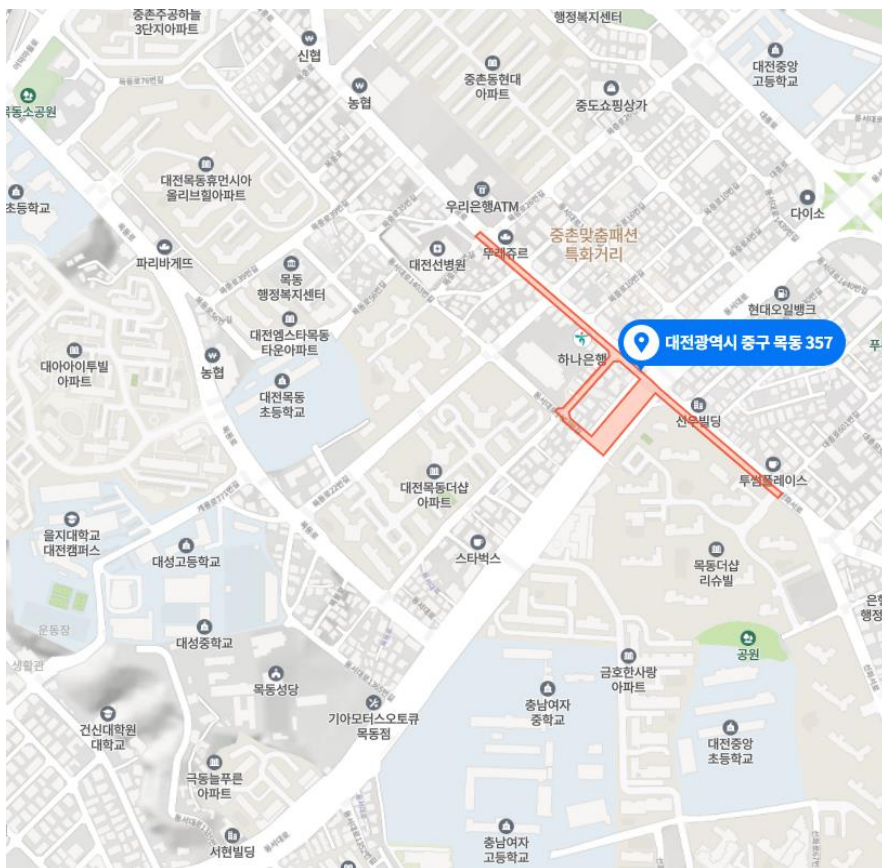


[사진 26] 대전광역시 중구 오류동 116-8

다음으로 2순위에 해당하는 횡단보도의 사진은 다음과 같다.



[사진 27] 대전광역시 중구 중촌동 418-1



[사진 28] 대전광역시 중구 중촌동 418-1

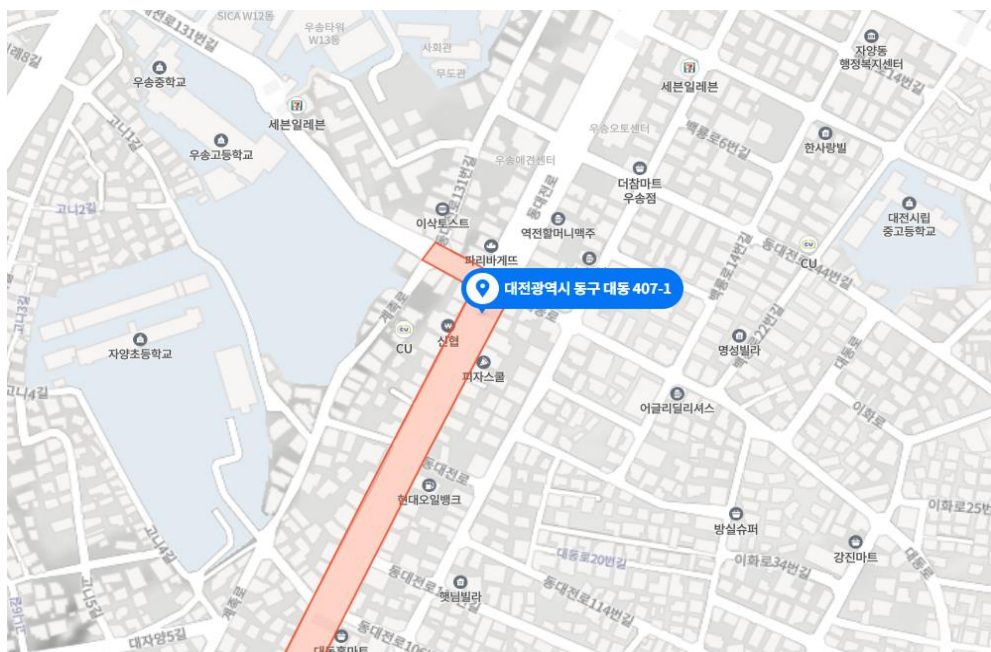
해당 지역은 주위에 학교, 아파트, 상가가 적절하게 밀집하여 균형을 이루고 있는 지역으로 횡단보도 위치는 학교 밀집지역에서 주거 밀집지역으로 넘어가는 사거리에 위치해있다. 따라서 해당 횡단보도에 스마트 횡단보도를 설치를 하게 된다면 학생과 주거 밀집 지역에 거주하고 있는 보행자 모두의 안전에 많은 도움이 될 것이라 판단된다.

#### <대전광역시 동구>

대전광역시 동구에서 1순위로 선정된 횡단보도는 다음과 같다.



[사진 29] 대전광역시 동구 대동 407-1



[사진 30] 대전광역시 동구 대도 407-1

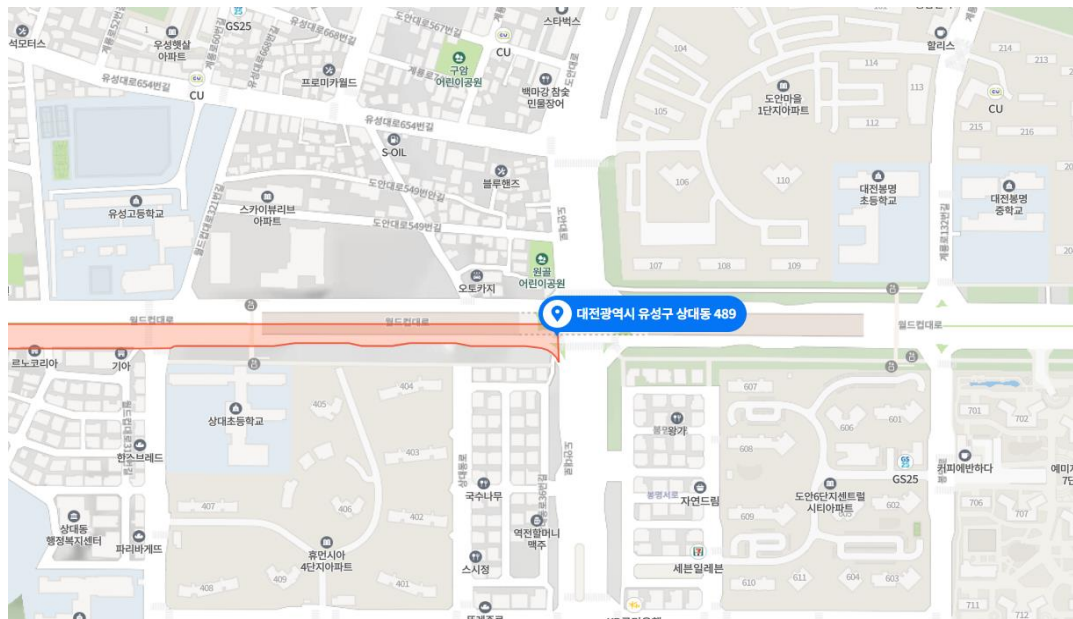
해당 횡단보도는 주위에 학교와 상가가 많은 지역으로 학교 밀집 지역에서 빌라 밀집 지역으로 넘어가는 곳에 위치한 횡단보도이다. 이 횡단보도 역시 위와 동일하게 스마트 횡단보도를 설치하기에 적합한 지역으로 판단된다.

### <대전광역시 유성구>

대전광역시 유성구에서 1순위로 선정된 횡단보도는 다음과 같다.



[사진 31] 대전광역시 유성구 봉명동 498



[사진 32] 대전광역시 유성구 봉명동 498

[사진 34] 대전광역시 대덕구 송촌동 520

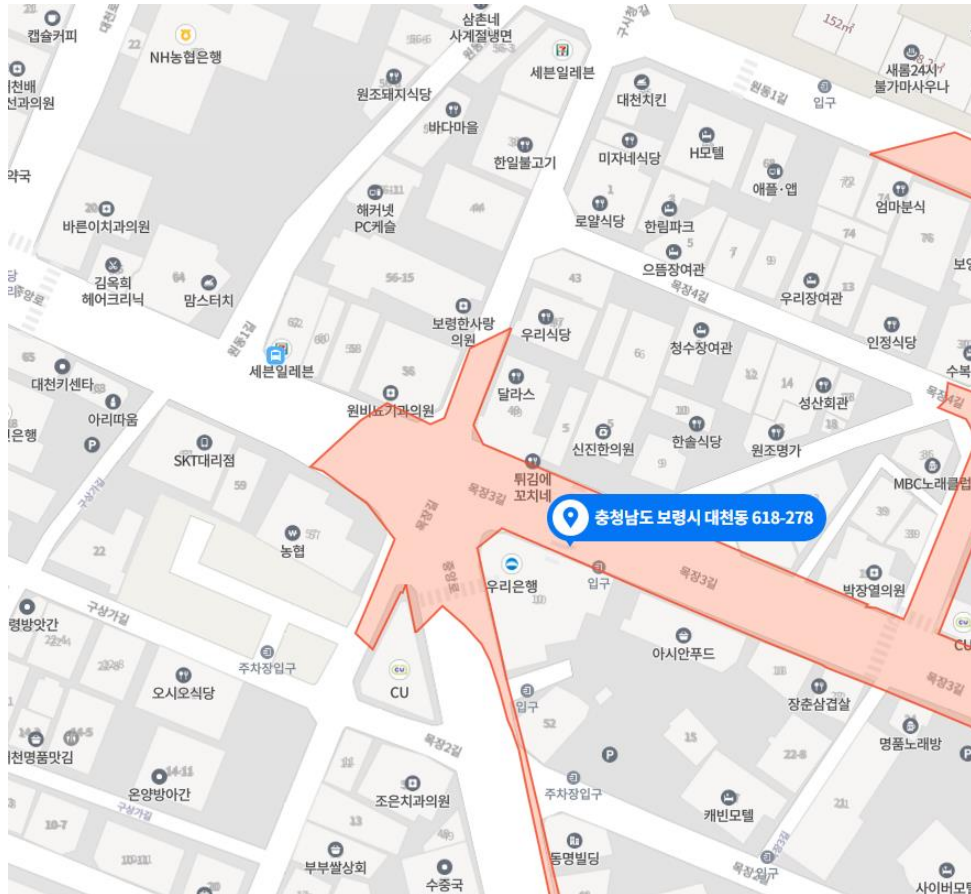
해당 횡단보도는 아파트, 빌라, 상가가 적절히 섞인 주거 복합 지역이다. 또한 차로수가 크고 교차로가 복잡하여 많은 교통량이 있다는 것을 알 수 있다. 따라서 운전자와 보행자 모두를 위한 스마트 횡단보도 서비스를 도입한다면 좀 더 안전한 보행이 가능할 것으로 예상된다.

#### <충청남도 보령시>

충청남도 보령시에서 2순위로 선정된 횡단보도는 다음과 같다.



[사진 35] 충청남도 보령시 대천동 618-278



[사진 36] 충청남도 보령시 대천동 618-278

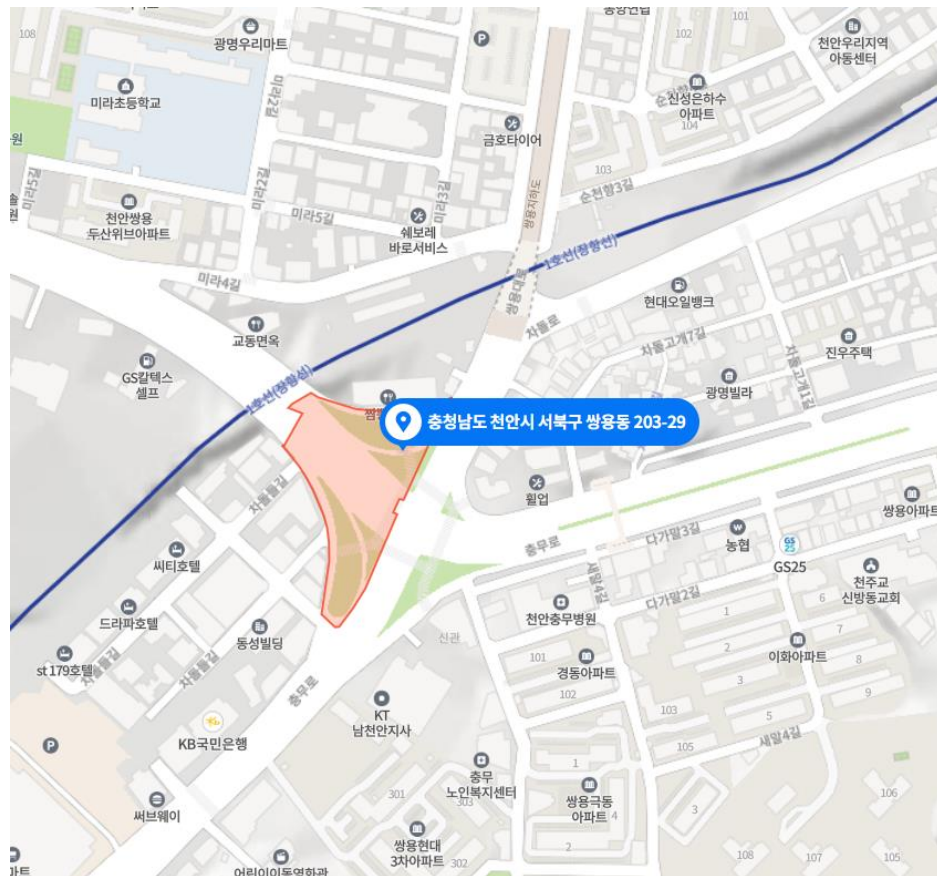
해당 횡단보도는 여러 도로들이 합쳐지는 지점에 위치하고 있으며, 차로수의 크기에 비해 교통량이 많아 혼잡도가 높은 지역이다. 따라서 스마트 횡단보도 서비스 중 'LED 횡단보도'와 같은 보행자를 위한 서비스가 아닌 운전자의 집중을 위한 서비스가 필요하다. 따라서 'LED전광판', 'IoT기반 과속방지 시스템'과 같은 서비스도 도입하여 운전자 주의 경고를 제공한다면 보행자 안전에 더욱 효과적일 것으로 예상된다.

#### <충청남도 천안시>

충청남도 보령시에서 2순위로 선정된 횡단보도는 다음과 같다.



[사진 37] 충청남도 천안시 서북구 쌍용동 203-29



[사진 38] 충청남도 천안시 서북구 쌍용동 203-29

해당 횡단보도는 아파트 단지과 빌라가 밀집한 지역에 위치해있다. 대전광역시의 대덕구와 비슷한 특징을 가지고 있기 때문에 추가 설명은 생략하겠다.

이처럼 클러스터링을 통해 1차로 횡단보도 중 각 지역에 대해 우선 선정 지표가 높은 스마트 횡단보도 설치를 위한 횡단보도 1개씩을 살펴보았다. 실제로 각 지역을 지도를 통해 살펴봄으로써 본 연구에서 제시한 '우선 선정 지표'가 타당하다는 것을

확인할 수 있었다. 또한 '충청남도 보령시'와 같이 각 지역의 특징에 따라 스마트 횡단보도 시스템 중 어떤 서비스가 해당 지역에 적합할 것인지도 제시하였다. 현재 제시된 예시를 제외하고도 '우선선정지표'를 활용한다면 스마트 횡단보도 설치가 우선적으로 필요한 지역을 쉽게 확인할 수 있을 것이다.

## b. 한계점 및 향후 발전 방향

본 연구는 횡단보도표준데이터 수집의 어려움, 주위 시설 데이터 수집의 어려움, 횡단보도 좌표를 통한 거리 계산의 한계, 신고 데이터의 설명 부족으로 인한 4가지 한계점을 가지고 있다.

첫번째로, 횡단보도표준데이터 수집의 어려움이다. 본 연구를 위해 충청남도, 대전광역시, 세종특별자치시에 대한 모든 횡단보도 데이터를 수집하고자 했지만, 충청남도의 경우 천안시, 서산시, 보령시, 태안군, 서천군, 예산군 6개 지역만 수집이 가능하였다. 횡단보도를 분석하지 못한 지역 중 충청남도의 아산시, 당진시의 경우 교통사고 관련 신고 건수, 아동 청소년 인구 비율, 학교 수, 인구 1명당 교통사고 발생 건수가 비교적 높은 편에 속한다. 또한 공주시의 경우 공주시의 경우 학교 수를 제외한 나머지 항목에서 모두 높은 편이며, 특히나 인구 1명당 교통사고 발생 건수가 매우 높은 편에 속한다. 따라서 충청남도 아산시, 당진시, 공주시의 횡단보도표준 데이터를 구할 수 있다면, 모델의 정확성과 보다 타당성을 더욱 높일 수 있을 것이라 예상되며, 교통사고가 잦다고 예상되는 아산시, 당진시, 공주시에도 '스마트 횡단보도'를 도입할 수 있는 근거를 제공할 수 있을 것이다.

두번째로는 주위 시설 수집의 어려움이다. 본 연구에서는 기존의 횡단보도표준데이터에 존재하는 열 뿐만 아니라 '학교수', '어린이시설수', '노인시설수', '아파트세대수', 'cctv수', '버스정류장수'와 같은 다양한 데이터를 병합하여 사고다발지역 주변 시설과의 상관성을 살펴보고자하였다. 하지만 여러 데이터를 병합하는 과정에서 횡단보도를 수집한 지역을 못 구하거나 구하더라도 데이터가 부정확한 경우가 많았다. 예시로 '노인시설수'와 같은 경우 전국 데이터를 사용하였지만 몇 지역에서는 아예 수집이 되지 않은 것을 확인할 수 있었다. 따라서 만약 횡단보도 데이터와 해당 지역에 대한 주변 시설에 대한 정확한 데이터를 통해 예측 모델링을 생성한다면 보다 정확한 피쳐 중요도를 얻을 수 있을 것으로

예상된다. 또한 클러스터링을 하는 과정 역시 현재보다 좀 더 다양한 지역에 고르게 선정될 수 있을 것이라 예상된다.

세번째는 횡단보도 좌표를 통한 거리 계산의 한계이다. 예시로 일부 교차로의 경우 4차선과 2차선이 함께 있는 경우도 존재한다. 따라서 4차선에서 교통사고가 발생하더라도 근처 1차선 횡단보도가 50m 내에 존재한다면 1차선임에도 교통사고건수가 높게 집계되기도 한다. 따라서 해당 횡단보도를 중복 데이터라고 판단하고 제거 하려하였지만 다양한 경우의 수가 존재하여 모두 고려하기 어려웠다. 따라서 만약 이러한 경우를 제외하고 피쳐 중요도를 판단한다면 차로 수의 피쳐 중요도가 더욱 올라갈 것으로 예상된다.

마지막은 신고 데이터에 대한 설명 부족이다. 본 연구의 목적은 단순히 교통사고 다발 지역을 선정하는 것이 아닌 어린이, 노약자와 같은 보행자 취약 횡단보도를 선정하는 것이다. 하지만 예측 모델링에 사용한 신고수와 같은 경우 보행자 사고 뿐만 아니라 차대 차 사고, 음주 운전 사고와 같은 다양한 교통 관련 신고가 혼재되어있다. 따라서 만약 보행자 관련 교통사고만으로 모델링과 클러스터링을 진행하게 된다면 보행자 교통 취약 지역을 더욱 정확하게 선정할 수 있을 것으로 기대된다.

이렇게 본 연구에서는 데이터 수집과 부정확성에 대한 어려움에도 사고다발지역에 영향을 미치는 변수들을 파악하고 해당 변수도의 중요도를 기반으로 스마트 횡단보도 입지 선정에 대한 '우선 선정 지표'를 제시를 하였다는 점에서 의의가 있다.

### **c. 연구를 통한 기대효과**

본 연구에서는 횡단보도에서 발생한 신고 데이터를 활용하여 사고다발지역을 지정하고 횡단보도의 사고다발 지역을 예측하는 분류 모델을 구축하였다. 앞서 구축한 모델에서 사고다발지역의 횡단보도 특징과 주변 시설에 대한 피쳐 중요도를 추출하여 각각의 변수가 횡단보도 사고에 미치는 영향의 정도를 파악하였다. 또한 상위 중요도를 가진 피쳐들 추출하고 클러스터링에 활용함으로써 보다 많은 교통

약자들이 도움을 받을 수 있는 횡단보도 취약 군집을 선정하였다. 추가적으로 주변 시설과 사고를 통합적으로 고려한 '우선선정지표'의 수식을 제시하였다.

따라서 이러한 객관적인 지표는 한정된 예산에서 스마트 신호등이 최우선적으로 설치되어 가장 효율적으로 보행자의 안전을 지킬 수 있는 방법을 제시하였다는 데에 그 의의가 있다. 또한 해당 '우선선정지표'는 본 연구에서 제시한 지역 외에도 범용적으로 사용이 가능하다. 이를 통해 본 연구는 각 지역에 맞는 스마트 횡단보도 도입 여부 결정에 기여할 수 있을 것으로 기대되며, 스마트 횡단보도 설치로 기대되는 효과는 다음과 같다.

'스마트 횡단보도'의 설치를 통해, 시민들의 안전한 통행권을 확보하고 교통사고를 예방함으로써 인명사고를 방지할 수 있다. 본 연구를 통해 선정된 최적 입지는 인근 어린이와 청소년, 노인과 장애인 시설이 밀집되어 있는 곳으로, 교통사고 취약 계층이 빈번히 이용하는 횡단보도이다. 이와 같은 교통약자들을 위해 LED 발광패널과 지향성 스피커 기능을 제공하여, 시각 및 청각적 무단횡단 및 사고를 예방하고자 한다.

스마트 횡단보도는 LED 전광판을 통해 정지선 위반에 대한 경고를 표시할 수 있기 때문에 보행자 뿐만 아니라 운전자들의 사고 또한 예방할 수 있다. 따라서 위에서 제시한 차로 수에 비해 도로가 복잡하고 교통량이 많은 '충청남도 보령시'의 최적입지와 같은 경우 'IoT기반 과속방지 시스템'과 같은 스마트 횡단보도 서비스가 도입하여 운전자 주의 경고를 제공한다면 보행자 안전에 더욱 효과적일 것으로 기대된다.

스마트 신호등과 같은 '사물인터넷(IoT)'을 통해 사고예방 뿐만 아니라 다양한 교통관련 데이터를 수집할 수 있으며, 차후 이를 통해 얻은 데이터들은 사고예방 뿐만 아니라 지역의 교통환경 개선에 있어서도 효율적인 정보를 지속적으로 공급하는 부가적인 효과 또한 기대할 수 있으며, 이러한 정보를 통한 분석을 본 연구와 같이 이용한다면, 더욱 안전하고 쾌적한 도심의 교통환경을 이루어낼 수 있을 것으로 예상된다.

## 8. 참고문헌 및 데이터 출처

이수현, 서용원, 김세인, 이재경 and 윤원주. (2022). 머신러닝을 활용한 어린이 스마트 횡단보도 최적입지 선정 - 창원시 사례를 중심으로 -. KIBIM Magazine, 12(2), 1-11.

<https://www.atnnews.co.kr/news/articleView.html?idxno=58046>

<https://safeupt.modoo.at/?link=aj8y1gwz>

[https://www.mois.go.kr/frt/bbs/type010/commonSelectBoardArticle.do?bbsId=BBSMSTR\\_000000000008&nttId=68567](https://www.mois.go.kr/frt/bbs/type010/commonSelectBoardArticle.do?bbsId=BBSMSTR_000000000008&nttId=68567)

공공데이터포털(<https://www.data.go.kr/>): 전국횡단보도표준데이터

공공데이터포털(<https://www.data.go.kr/>): 전국노인장애인보호구역표준데이터

공공데이터포털(<https://www.data.go.kr/>): 전국어린이보호구역표준데이터

공공데이터포털(<https://www.data.go.kr/>): 전국 버스 정류장 위치 정보

공공데이터포털(<https://www.data.go.kr/>): 전국무인교통단속카메라표준데이터

k-apt 공동주택관리 정보시스템: k-apt관리비공개의무단지 기본정보

kosis: 행정구역별 지목별 국토이용현황

행정안전부: 연령별 인구현황

교통사고정보 개방시스템: 보행자 사고다발지역정보 데이터셋

GIS DEVELOPER, 대한민국 최신 행정구역(SHP), (<http://www.gisdeveloper.co.kr/?p=2332> )